

RADIO ENGINEERING and ELECTRONIC PHYSICS

English Edition of

РАДИОТЕХНИКА И ЭЛЕКТРОНИКА

Published by the American Institute of Electrical Engineers
with the aid of a grant from the National Science Foundation

No.6 June 1961

Translated and Produced by Royer and Roger, Inc.

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Established 1884

345 East Forty-Seventh Street
New York 17, N. Y.

Warren H. Chase, President
N.S. Hibshman, Executive Secretary
C.E. Dean, Technical Vice President, Communications
W.F. Denkhau, Director of Publications
L.G. Abraham, Chairman, Communications Division

The English edition of RADIO ENGINEERING AND ELECTRONIC PHYSICS is published by the American Institute of Electrical Engineers with the aid of a grant from the National Science Foundation. © 1961 by American Institute of Electrical Engineers. Also published under the same arrangement are the Russian electronic journals RADIO ENGINEERING and TELECOMMUNICATIONS.

RADIO ENGINEERING AND ELECTRONIC PHYSICS

(РАДИОТЕХНИКА И ЭЛЕКТРОНИКА)

Publication of the Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

Translated and Produced
by
Royer and Roger, Inc.



Translation Editor: Ivo Herzer, Columbia University

AIEE REVIEW COMMITTEE FOR RADIO ENGINEERING AND ELECTRONIC PHYSICS

Leonard S. Schwartz
New York University College of Engineering
Chairman

A.W. Bickley	A. Burr Fontaine	W. Miller	C.A. Stutt
W.P. Birkemeier	F.E. Froelich	Harry Rowe Mimno	G.C. Sziklai
T.T.W. Bucher	Paul H. Gleichauf	W.W. Peterson	Joseph Vogelmann
J.L. Callahan	G.S. Glinski	B. Reiffen	G.M. White
G.R. Cooper	Bernard Harris	W.G. Schmidt	F.B. Wood
W.A. Depp	R.K. Hellmann	Herbert Sherman	H.L. Yudkin
R.S. Enticknap	D.E. Higginbotham	D.L. Solomon	

Subscriptions to Radio Engineering and Electronic Physics should be sent to AIEE
Special Subscription Department

41 East 28th Street, New York 16, New York

1961 Subscription rates:

	\$	£
Individuals	28.50	10
Libraries, institutes, govt. agencies	57.00	20
12 issues per annum comprising approximately 1900 pages		

LONG-RANGE TROPOSPHERIC PROPAGATION OF ULTRASHORT WAVES (SURVEY)

N.A. Armand, B.A. Vvedenskiy, A.I. Kalinin, M.A. Kolosov,
A.V. Sokolov, A.V. Shabelnikov, R.A. Shirey

An analysis of experimental and theoretical work on long-range tropospheric propagation of ultrashort waves is given.

The field intensity is examined as a function of distance, wavelength and time. The gain losses of antennas, the distortion of the signal, and the band of transmitted frequencies are all analyzed.

Various theories describing the mechanism of long-range tropospheric propagation of radio waves are considered. On the basis of the comparison with experimental data, it is shown that not one of the existing theories can completely explain all the available experimental results.

I. REVIEW OF EXPERIMENTAL WORK*

INTRODUCTION

The large number of fundamental experimental studies of long-range tropospheric propagation of ultrashort waves conducted in different countries has added considerably to the extent in depth of our knowledge of the characteristic features of this phenomenon.

From the considerable amount of experimental material, it has been concluded that the region of ultrashort waves can be used for high-quality transmission and various types of information (telephone conversations, broadcasting, and isolated experiments on television transmission), on radio relay systems with the intermediate stations at a distance of 600 km from one another. This distance is by no means a limit in tropospheric lengths. Work on the practical application of long-range tropospheric propagation is being conducted in many countries.

Article [1] describes the results of an experiment on the reception and transmission of signals through long-range propagation over distances greater than 1300 km. In the experiment, transmitters with a wavelength of 75 cm and a power of 50 kw, sensitive narrowband receivers, and parabolic antennas 18 m in diameter were used. According to the data of [23], very weak signals were detected at a distance of 1600 km on a wavelength of 1.4 m.

These successful achievements were made possible as a result of the improvement of the transmitting-receiving apparatus and antennas, the use of auxiliary instruments greatly facilitating the analysis of experimental data, and a statistical approach to the analysis and study of the experimental data. The development of communications, broadcasting and television requires further studies on the dependence of the field on the distance D , the antenna characteristics, the path parameters, the meteorological conditions, and the distinguishing climatic

*This article consists chiefly of a paper submitted to the Soviet National Committee of the International Scientific Union in May, 1960. The paper was presented at the Twelfth General Assembly of the International Scientific Union held in September, 1960 in London.

features. This is also important in establishing the true physical nature of the phenomenon and in clearly evaluating the role of the mechanisms of scattering, reflection, and diffraction. Considerable attention should be devoted to a question of practical importance — the investigation of the bandwidth of the transmitted frequencies at various distances and the study of the possible distortions of the information, the phase stability, etc.

Since these questions are completely covered in many works, [1-13], [37, 32, 43], we believe that we may concentrate our attention on the difficult problem of comparing and analyzing the existing measurement data, which has not been fully systematized and is, in part, contradictory. The interpretation given this material is sometimes biased, a circumstance which is aggravated by our lack of knowledge concerning the structure of the troposphere.

1. DEPENDENCE OF THE FIELD ON DISTANCE

The received power P_R generally comes from the radiation of the electromagnetic energy by different tropospheric inhomogeneities. Because of changes in the tropospheric inhomogeneities, this power is a random function of time. Therefore the function $P_R(D)$ can be determined more or less definitely only as an average over fairly long periods of time.

A comparison of functions $P_R(D)$ obtained under different climatic conditions, or, in different seasons, is difficult, especially because of differences in duct parameters, their surface, antenna heights, local terrain, and the total duration of the measurements.

Because of the many factors affecting long-range propagation, the opinion was even expressed, [57] that it was impossible to accurately predict the attenuation in specific ducts without conducting fairly prolonged measurements in these ducts.

In attempting to obtain some general picture of the phenomenon in the range $\lambda = 3$ to 300 cm, we analyzed prolonged measurements of the quantity P_R/P_0 (P_0 is the value of the power in free space) made over the past four to five years. For the greater part, these measurements were conducted over fixed distances near to the earth's surface, and to a smaller extent, on apparatus set up on airplanes, ships and trucks.

In [15, 16, 17, 21, 23], as in the results of Megow, [53], local maxima of P_R/P_0 were found. In isolated instances, these local maxima were clearly the result of interference. These results can be regarded as proof that radio waves are reflected from layers in the atmosphere.

In a number of measurements of $P_R(D)$, different propagation theories were verified experimentally, [15, 14] and data for different climatic conditions were compared. It was found, (for instance in [16]) that when the computed value of the scattered power coincided with the experimental value $\lambda = 9$ cm, the average level of P_R/P_0 at $\lambda = 3$ cm was 15 db lower than the computed value. Moreover, in comparing the values of P_R obtained in England at

frequencies of 3,480 and 858 Mc [19, 33] with the forecasting curve constructed by Norton and others, [31], which is characteristic of conditions in the USA and is based on extensive experimental material, it was found that the P_R given by this curve were in close agreement only for 3480 Mc and that they were too low for 858 Mc. The effect of climatic conditions cannot be ignored here. Other indications of this such as those in [13], make it possible to assert that the results obtained in England do not hold unconditionally for other climatic regions. This factor was not taken carefully into account in the nomograms [35], since they were obtained by simply averaging the measurements obtained for different λ and different climatic regions. A comparison of the data for the function $P_R(D)$ obtained by Bullington, [36] and Norton and others [54] also shows a mutual discrepancy.

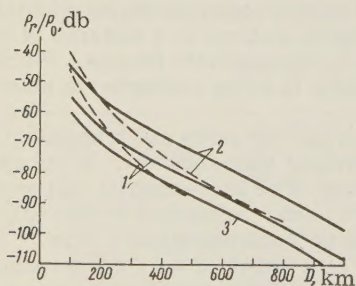


Figure 1. A comparison of the dependence of the field on distance according to the experimental data of Norton (solid lines) and Bullington (dotted lines):

1) $\lambda = 30$ cm; 2) $\lambda = 3$ m; 3) $\lambda = 10$ cm.

To determine the influence of the wavelength and the climatic conditions and the function $P_R(D)$, we carefully sorted the results of [1, 5, 10, 16, 17, 21, 23, 24, 25, 26, 28, 29, 33, 32, 118] in terms of λ and analyzed them (Figure 2). Despite the considerable spread of the points — the results of both random errors in the measurements, and the difference in the duration and season (or day) of the measurement — it can be seen that, for instance, the values for England (marine climate) are somewhat higher than the values for the USA.

2. DEPENDENCE OF THE SIGNAL LEVEL ON FREQUENCY

The dependence of the signal level on frequency has not been studied sufficiently. The different theories give a very different picture (from $P_R \sim \lambda^{1/3}$ to $P_R \sim \lambda^2$, and from a constant linear attenuation α to $\alpha \sim \lambda^{-1/3}$). The majority of experimental studies show a decrease of P_R/P_0 with increasing frequency F (Figure 2), with the P_R/P_0 being 20 db lower for $\lambda = 3.2$ cm than for $\lambda = 10$ cm.

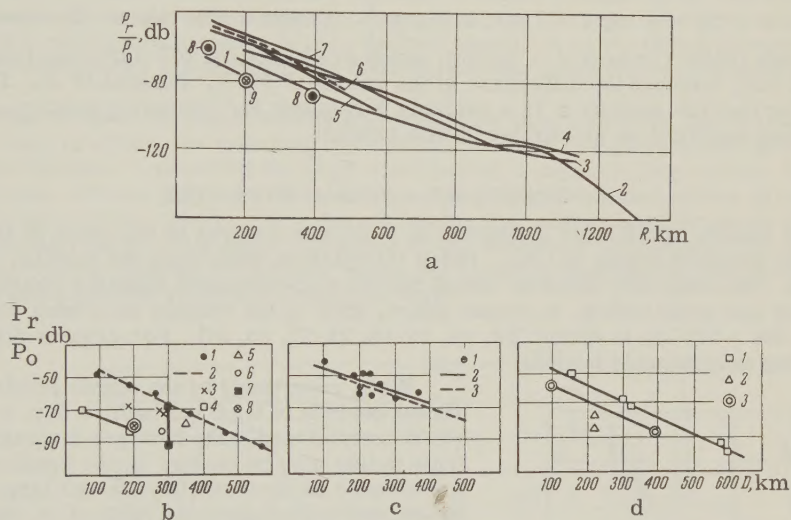


Figure 2. Dependence of the field on distance according to data of different experiment studies.

a: 1) [16], $\lambda = 3$ cm, 2) [21], $\lambda = 75$ cm, 3) [10], $\lambda = 75$ cm, 4) [25], $\lambda = 3.5 - 1$ m and shorter, 5) [23], $\lambda = 35$ cm, 6) [17], $\lambda = 10$ cm, 7) [24], $\lambda = 3.3$ m, 8) $\lambda = 30$ cm, 9) $\lambda = 8$ cm; b: 1) [17], $\lambda = 10$ cm, 2) approximation [17], 3) [5], $\lambda = 7.4$ cm, 4) [1], $\lambda = 100 - 400$ km, and $f = 100$ and 4000 Mc; and according to [56], the relative attenuation is 20 db for $D = 175$ km, and $f = 100$ and 1300 Mc. Study [5, 38] also indicate that there are differences in the functions $P_R(\lambda)$ when the average is taken over different time intervals for different times of the day and different distances.

However, so far the experimental studies have shown no similarity in the function $P_R(\lambda)$. Thus, according to [40], the relative attenuation with increasing frequency is 14 db for $D = 300$ km, and $f = 100$ and 4000 Mc; according to [57], the relative attenuation is 9 db for $D = 100 - 400$ km, and $f = 500$ and 5000 Mc; and according to [56], the relative attenuation is 20 db for $D = 175$ km, and $f = 100$ and 1300 Mc. Study [5, 38] also indicate that there are differences in the functions $P_R(\lambda)$ when the average is taken over different time intervals for different times of the day and different distances.

It is natural to see in this the effect of the changeability of the structural conditions of the troposphere and of the meteorological conditions. Therefore, the use of the statistical approach to evaluate the function $P_R(\lambda)$ in work [38] should be regarded as an achievement.

Here the function was determined under winter conditions in a duct 96 km long with simultaneous measurements of 417 and 2290 Mc. It was possible to show that the ratio of the mean hourly values of the signals at the two frequencies was distributed about a standard logarithmic curve and that during the entire period of observation the nature of the variation of the ratio as a function of λ changed within the limits of λ^{-2} and λ^2 . For 50% of the time this ratio was proportional to λ . As known, the linear attenuation α of the average value of $P_R(D)$ beyond the horizon is considerably smaller than within the region of "classical" diffraction. However, there are relatively few quantitative data on direct measurement of α . There is practically no data on the function $\alpha(D)$, and its relation to the season and time of day, with the exception of [1, 21, 25], where the data on the function $\alpha(D)$ are contradictory. Likewise the results of a number of works, [1, 16, 17, 21, 25, 58] permit the representation of $P_R(D)$ in the form of an exponential function indicating that α is independent of D (the curve $\log P_R$ vs. D is a straight line). With regard to the constancy of $\alpha(D)$, see also [57].

Figure 3 gives α as a function of λ . We constructed this curve on the basis of a statistical processing of the measurements, [1, 16, 17, 21, 23, 24, 33, 39]. The average values of $\alpha(\lambda)$ for different climatic regions can be satisfactorily approximated on the curve $(\alpha, \log \lambda)$ by a straight line $\alpha = C_1 - C_2 \log \lambda$. Thus, $\alpha \sim \log \frac{1}{\lambda}$. Figure 3 also shows the result of our measurements (point 1, where $\lambda = 30$ cm), which are in almost full agreement with the data in [18] ($\lambda = 35$ cm), although the difference in the values of P_R/P_0 attained 10 db. There is basis for supposing that the quantity α is a variable parameter for comparing results and for determining certain quantitative properties of this process.

3. CHANGE OF A SIGNAL WITH TIME

A large amount of attention is devoted in a number of works to the study of fluctuations or fading on the received signal in time. These fluctuations determine the stability of communication links. The extensive material which has been accumulated makes it possible to distinguish slow and rapid fading. In recent years, most of the studies have been conducted for ducts up to 300 - 500 km in extent [13, 14, 16-19, 21-29, 33, 57]. For greater distances, the data on fading is extremely limited.

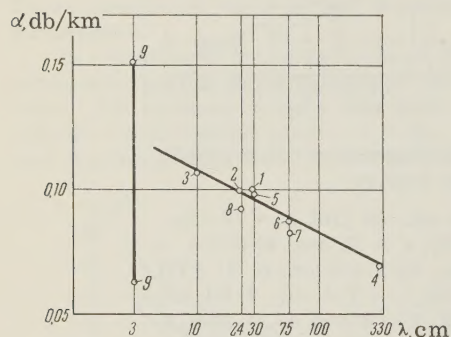


Figure 3. Dependence of linear attenuation α on the wavelength λ .
1) $\lambda = 30$ cm; 2) 39; 3) 17; 4) 24; 5) 33; 6) 21; 7) 1; 8) 23; 9) 16.

When recordings of the signal levels over different periods of time (an hour, a day, a month, a year) are analyzed, slow fading is invariably observed. This fading may be caused by the appearance or disappearance of inversion layers and large inhomogeneities, and by changes in the value of $d\epsilon/dh$ (the vertical gradient of the permittivity of air ϵ).

The diurnal behavior of the mean hourly values of the signal is not expressed clearly and does not have any definite pattern [26], although most frequently the signal is stronger in the evening and at night than in the daytime [28, 41, 56, 59]. In summer, this behavior is seen more clearly than in winter, and it is more definite over short distances (100 - 150 km) than over long distances (400 - 500 km). In works [33], it is noted that at distances of 158 and 332 km, the signal is stronger at night and in the evening than by day; at a distance of 575 km, this behavior is reversed and the maximum values of the signal are observed in the daytime hours. According to our measurements on ducts of 100 and 390 km, the diurnal variation amounted

to 8 - 12 db.

The amplitude of the slow fading does not seem to depend on frequency. Under different climatic conditions, the difference attains 10 - 25 db (distances of 300 - 400 km) [19, 24, 25, 57], sometimes attaining 300 db and more, and is greater in summer than in winter.

The level of the received signal also undergoes seasonal fluctuations, increasing in summer and decreasing in winter. The amplitude of the seasonal fluctuations decreases with

increasing distance. It is possible to judge of the degree of these changes in amplitude from work [25], according to which the difference between the maximum and minimum mean monthly values amounted to 12 and 6 db for distances of 300 and 987 km respectively.

During short periods of time, rapid fading is also observed. This is superimposed on the slow fading and, by contrast to the slow fading, depends on frequency. As a result of the superimposition of the rapid and slow fading, the total amplitude of the fading is increased. According to some sources, this amplitude attains a maximum at distances of 100 - 300 km [13, 28], but according to the results of other studies [40, 48], it decreases monotonically with increasing distance and, according to [25], equals 3 db in winter and 10 db in summer at a distance of 990 km.

In a large part of the studies, the distributions for short time intervals (tens of seconds to one hour) are close to Rayleigh distributions [21, 28, 42]. There is evidence that a Rayleigh distribution is encountered only for weak signals [19, 25]. However, other studies indicate that the amplitude of fading for short time intervals is often smaller and sometimes greater than 13.4 db [26, 59]. On the basis of measurements, a number of authors note that as the period of observation increases the distribution becomes a standard logarithmic one [25, 26, 42].

Experimentally it has been found [52] that patterns of the type described in [32] indicate that there is a constant component in the received signal. In cases where inversion layers or atmospheric waveguides are formed, the distribution may not be of the standard logarithmic type.

A detailed study of the distributions was conducted in articles [14 and 16]. As a result it was shown that, in effect, the following types of distribution are encountered in addition to Rayleigh distribution: a modified Rayleigh distribution, a standard logarithmic distribution, and other types. Article [16] expresses the hypothesis that the distributions in the first and third categories are the results of incoherent scattering, while distributions in the second category are the result of the superimposition of the scattered components and the constant components. This hypothesis is dubious, since the Rayleigh distribution can be obtained when radio waves are reflected from broken layers [44].

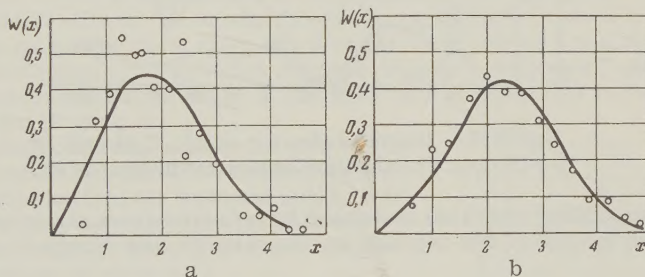


Figure 4. A comparison of the experimental curves of the distribution of the field amplitude with the theoretical curves.

A) Summer, $D = 150$ km, $\beta = 1.12$; B) winter, $D = 390$ km, $\beta = 1.5$.

From our observations on a duct 390 km in extent ($\lambda = 30$ cm), the integral curves obtained for periods of an hour or more were close to distributions of the standard logarithmic type. This points to the existence of a constant component. Even during minute intervals of time, a Rayleigh distribution was rarely encountered and then only for the weakest signal levels. It is possible to draw the conclusion regarding the existence of a constant component by comparing the experimental data with the probability density of the process depending on the constant component and the finite number of components subject to a Rayleigh distribution law.

Figure 4 shows a wintertime and summertime comparison of the theoretical probability density curves with the distribution parameters found from direct measurements. This shows that the equation

$$W(x) = xe^{-\frac{R^2 + a_0^2}{\Sigma a_s^2}} J_0 \left(\sqrt{\frac{2a_0^2}{\Sigma a_s^2}} x \right),$$

where

$$x = \sqrt{\frac{2}{\Sigma a_s^2}} R;$$

R is the total amplitude of the field vector, a_0 is its constant component, and a_s is its random component, can describe the changes of the signal level in time.

Figure 5 shows the diurnal behavior of the component a_0^2 and Σa_s^2 , and the ratio β^2 for a 150 km duct during one day. By contrast to the random components, the constant component is maximum in the evening and morning, and close to zero at about midday when the distribution is close to Rayleigh.

The fading rate — the number of times the instantaneous values of the signal intersect the average level in a second — will obviously be determined by the rate of change in the phase of each interfering wave in time. This rate should increase with increasing frequency and increasing duct length, as is confirmed by a number of experimental studies. A connection between the fading rate and the magnitude of the received signal is also seen [33]. An increase of the signal usually leads to a decrease of the fading rate.

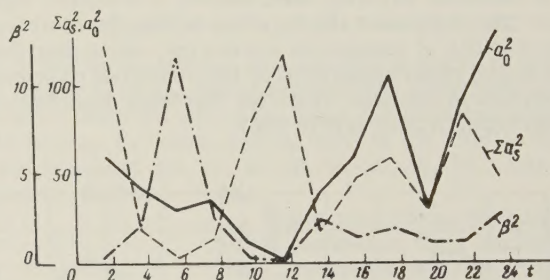


Figure 5. Diurnal behavior of a_0^2 , Σa_s^2 and β^2 . $D=150$ km. The abscissa indicates the time of day.

Rapid and deep fading may also be caused by reflection from airplanes. This type of fading increases with frequency and depends on the velocity and direction of motion of the airplane.

4. GAIN LOSSES IN ANTENNAS

In experiments extending far beyond the limits of the horizon, the phenomenon of gain losses in the antennas ("losses in the coupling of the aperture to the medium") were discovered. These losses were caused entirely by the special properties of propagation in the troposphere.

The essence of this phenomenon is that in antennas with a gain G exceeding 25 - 40 db the gain is not fully achieved and is smaller than in free space.

In the literature, data from experimental studies of this phenomenon beyond the limits of the horizon appeared in 1955 [2, 45, 48].

Although the mechanism of how the gain losses arise in antennas is far from completely studied, at least two hypotheses can be proposed on the basis of the experimental data to explain this phenomenon.

1. The propagation of radio waves in a statistically inhomogeneous medium ultimately leads to a distortion of the wave front in the plane of the receiving antenna and thus the energy absorbed by the receiving antenna is smaller than the energy it would absorb in the absence of fluctuations in the amplitude of phase.

2. Elementary waves with different random angles of incidence can arise at the receiving antenna. When the fluctuation of the angles of incidence are comparable with the width of the radiation pattern, the effective gain of the antennas is reduced.*

Both of the above mechanisms were discovered experimentally in [20] as a result of rapid oscillation of the radiation pattern of an antenna. The instantaneous characteristic of the radiation patterns were basically of two types: a) the width was $1.5 - 3\times$ broader than the width of the radiation pattern of the antenna taken within the limits of direct visibility; or b) the instantaneous width was not broader, but the direction of the maximum of the radiation pattern of the antenna relative to the axis of the duct changed in time.

The question of the first mechanism was studied in [51] with certain assumptions about the correlation function.

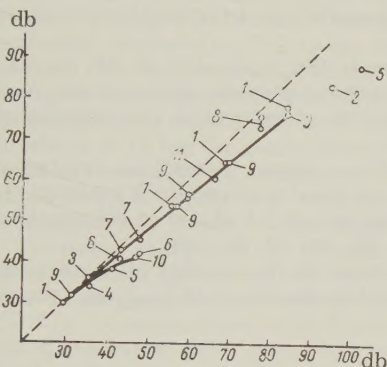


Figure 6. The actual gain of an antenna as a function of the gain for a plane wave (the abscissa shows the gain of the antenna for a plane wave, while the ordinate shows the actual gain).

1—[46], $D = 300$ km, $\lambda = 12.0$ cm; 2—[5], $D = 300$ km, $\lambda = 7.3$ cm; 3—[17], $D = 370$ km, $\lambda = 10.0$ cm; 4—[19], $D = 280$ km, $\lambda = 8.7$ cm; 5—[22], $D = 320$ km, $\lambda = 4.0$ cm; 6—[47], $D = 215$ km, $\lambda = 3.3$ cm; 7—[47], $D = 150$ km, $\lambda = 3.3$ cm; 8—[41], $D = 278$ km, $\lambda = 9.8$ cm; 9—[25], $D = 300$ km, $\lambda = 13.0$ cm; 10—[48], $D = 320$ km, $\lambda = \dots$; 11—[49], $D = 74$ km, $\lambda = 3.3$ cm.

Independently of [20], a similar hypothesis that both mechanisms could exist was expressed and theoretically studied in [55]. This phenomenon was studied experimentally in [5, 17, 19, 22, 25, 46, 47, 48] by a comparison of signals received at antennas with different gains. A comparison of the results of the measurements is made difficult by differences in the time and duration of the measurements, since in all cases these losses are given as averages, which, as a rule, are taken over different intervals of time. With small exceptions, all these measurements were conducted on ducts of about 300 km in extent and in the range of wavelengths of 3 to 12 cm (Figure 6).

Although the justification for adding the gain of the transmitting antenna to the gain of the receiving antenna is an open question (see for instance [104]), it is seen from Figure 6 that the experimental points are grouped with a small spread about a straight line approximating the results of work [25].

For ducts with an extent of the order 300 km, the gain losses increase as the total gain of the receiving and transmitting antennas increases and as D increases to 300 - 500 km, with $f = 2290$ Mc. However, there is evidence that with a further increase in distance these losses begin to drop. For instance, for distances of the order of 1000 km, when $f = 413$ Mc, all of the gain of the parabolic antenna 18 m in diameter is used [25] (see also [1]).

There is an almost total lack of data on the variations of the gain losses of antennas with seasons, the time of day, and meteorological conditions. However, from [5] it can be concluded that the magnitude of the

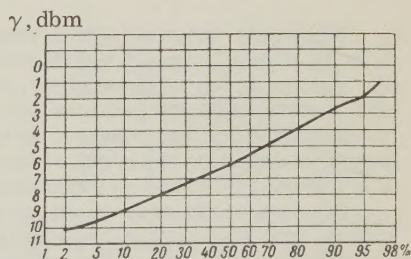


Figure 7. Distributions of the gain difference between two antennas. On the ordinate, γ indicates the difference between the actual gain of the antennas in dbm; the abscissa indicates the amount of time in % during which γ is greater than or equal to the corresponding ordinate.

*Because of fluctuations in the angles of incidence [49, 50], gain losses are sometimes discovered in the zone of direct visibility as well.

losses is statistical in nature, as is the magnitude of the received signal. On this basis a curve (Figure 7) can be constructed which definitely points to the existence of such a dependence. Article [5] contains values of the weekly averages of the power of signals received during a year and two antennas with a gain of 38 and 56 db. During the different seasons the quantity $\gamma = G_{18M} - G_{8.5M}$ changed from zero to 10 db. It was found that γ is larger for stronger signals than for weaker signals.

5. DISTORTION OF SIGNALS IN LONG-RANGE TROPOSPHERIC PROPAGATION. ATTAINABLE FREQUENCY BANDWIDTH

The arrival of many waves with different amplitudes and different time lags at the receiver leads to distortion of the signals and limits the possibility of transmitting messages with a fairly large bandwidth.

Experimental study of signal distortions are relatively few in number [5, 33, 42, 51, 61]. Tidd [62] has studied the possibility of transmitting broadband signals such as those needed for television. However, the study does not make it possible to determine the stability of transmission.

The published theoretical works on the determination of signal distortions are based on the representation of the troposphere in the form of some ideal quadrupole [63-66] or on a determination of the amplitude correlation of the components of the signal at different frequencies within the limits of the transmitted spectrum [30, 60, 67, 68, 69].

For antennas with a small directivity, all studies based on the concept of incoherent scattering lead to a relation between the maximum possible passband $(\Delta f)_{\max}$ and the distance D in the form

$$(\Delta f)_{\max} = \frac{\text{const}}{D^3},$$

which differs only in the value of the constant in different papers.

Under the assumption of coherent scattering [66], the following relation is obtained:

$$(\Delta f)_{\max} = \frac{\text{const}}{D}.$$

While with antennas with a small directivity the amplitude of waves having a large time delay is reduced through diffraction attenuation by the ground (which, it is true, is not taken into account in any of the studies), and through the "directivity" of the troposphere, in antennas with narrow radiation patterns the amplitude of these lagging waves is also reduced through the directivity of the antennas. In this case, the maximum passband will depend to a considerable extent on the width of the radiation pattern of the antennas. However, the nature of this dependence shows up differently in the literature.

Because of the random nature of the re-emission of electromagnetic energy by tropospheric inhomogeneities, the signal distortion should also be random. This is confirmed by measurements [5, 61] and experiments conducted in the USSR. The experiments show that both the amplitude distortions of the signal and the time delay are subject to large and rapid random changes in a passband several megacycles in width. The signal distortions of two antennas at a sufficient distance from one another are uncorrelated. Therefore diversity reception not only improves the signal level, but also reduces the distortions of the signal, and that provides for a large carrying capacity in tropospheric radio ducts.

Without doubt, both the experimental and the theoretical studies connected with the determination of signal distortion are far from complete. Further studies are necessary for the purpose of finding more carefully the dependence of the distortions and distance, wavelength, antenna directivity, and so on. It is also necessary to determine the effects of diversity reception on signal distortion. It seems advisable to develop theoretical studies of signal distortion by finding the correlation between the amplitude and phases of the components of the signal spectrum. This approach appears to be the most promising, since distortions can be determined statistically by this method.

6. RADIO-METEOROLOGICAL STUDIES

An analysis of the connection between the behavior of the signal and the meteorological conditions in the ducts may be of assistance in determining the predominant role of one or another mechanism of long wave propagation.

Refraction measurements [37, 70, 71] have added to the extent and depth of our conception of the structure of the troposphere and, in particular, to a knowledge of the function $\epsilon(h)$, $(\Delta\epsilon)^2$ and the size of the turbulences l . These measurements have shown [71, 72] that the $\sqrt{(\Delta\epsilon)^2}$ usually varies within the limits of 0.3 - 3 N units and there regularly exist inhomogeneities from 1 to 300 m thick with layers in them. The intensity of the "jumps" in these layers varies from 2 to 50 - 60 N units and is especially great in the so-called "invisible" clouds [39, 72]. The existence of turbulent layers has also been established.

In article [72], it is asserted that at height $h = 3000$ m and more the quantity $(\Delta\epsilon)^2/l$ is too small to explain the distant field and that the change in the quantity with height does not give the required function $P_r(D)$. No large inhomogeneities of ϵ were found in the troposphere, while at altitudes $h \cong 15$ km the fluctuations of ϵ lay below the threshold of sensitivity of apparatus. The experiments conducted in the USSR up to an altitude of 5 km [70] did not show a decrease of $(\Delta\epsilon)^2$ with height. The intensity of $(\Delta\epsilon)^2$ increased sharply at the boundary of inversions and clouds and was not connected with the values of ϵ near the ground. We conducted simultaneous measurements of P_r and $(\Delta\epsilon)^2/l$. A comparison with Gordon's formula [64] (Figure 8) shows that for a duct 300 km in extent the computed value of P_r is 30 db higher than the experimental value.

A radio method can also be used to obtain information on the structure of the atmosphere and to determine the mechanism of long-range propagation. For instance, a study of the characteristics of the radio signal made it possible to study the drift of inhomogeneities and to show that inversion layers have an appreciable influence on wavelength in the meter range in ducts 100 - 400 km in extent [73, 74].

A large amount of attention is devoted to determination of the variation of P_r with a gradient of refractive index of air ΔN_1 in the 0 - 1000 m layers. Since a clear relationship has been established [75, 76] between ΔN_1 and the quantity $N = N_0$ at the surface of the earth, some authors [27, 77] use the quantity N_0 to predict the average values of P_r . For instance, in [77] Bean established that when the quantities P_r and N_0 are averaged over a week or a month the correlation coefficient r between them is equal to 0.8 - 0.95. The correspondence coefficient k which he computed per 1 N unit turned out to be equal to 0.2 db/1 N units on the average. On the basis of the quantity k and world maps of isolines of N_0 , it is shown in [78] that the change of P_r in accordance with climate attains 32 db.

However, the parameter N_0 , which characterizes fairly well the average gradient dN/dh and less well the intensity of the jumps of ϵ at the boundaries of atmospheric inhomogeneities and inversion layers, gives no information on the height of the layers or on the dimensions of turbulent formations. The quantities r and k show diurnal variation, having a maximum in the early morning hours and a minimum in the afternoon. Changes of the average values of k are also observed in different climatic regions. Thus in [39] the average value of k turned out to be equal to 0.8 db/1N unit (Italy), $D = 400$ km.

In regard to a study of the relationship between ΔN_1 and N_0 , [59] notes that no connection between these quantities was found in the eastern part of Africa, (Dakar), Australia, and the Cocoa Islands.

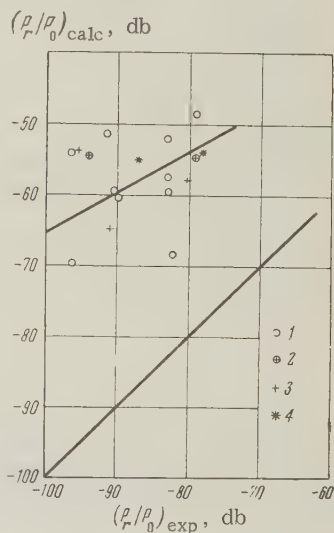


Figure 8. A comparison of the experimental values of the field intensity with the computed values.

- 1) $\lambda = 140$ cm; 2) $\lambda = 100$ cm;
- 3) $\lambda = 10$ cm; 4) $\lambda = 30$ cm.

Misme [80] does not believe that the quantity N_0 and ΔN_1 take into account the influence of the upper layers of the atmosphere and introduces a new meteorological parameter G_M . This parameter is based on a calculation of the gradient N in the layer from the ground to the region of intersection of the radiation pattern of the antenna, as well as on the thermodynamic stability of the atmosphere. In analyzing the connection between P_r and G_M , Misme finds a fairly clear relationship between these quantities. However, it is difficult to form a judgement on the advisability of using this parameter because of the lack of the comparison of P_r with ΔN_1 and N_0 under the same conditions in [80]. The connection of P_r with general meteorological conditions is also studied here.

Klinker [81] has established that in anticyclonic conditions, which are characterized by sinking depressions, the signal undergoes slow fading of 10 - 20 db with a period of 10 - 30 min. Its magnitude is 10 - 20 db higher than the average value.

On the contrary, in cyclonic conditions, which are characterized by intensive stirring of the atmosphere, the signal undergoes rapid fading after 30 db with a period from 1 sec to a few minutes, while its magnitude is 20 db lower than the average value. These signal characteristics are in agreement with [82], in which fading is studied in connection with the passage of frontal surfaces. There it is shown that the movement of a front along a duct combined with atmospheric instability near the frontal surface leads to rapid and deep fading of the signal.

The measurements described in [33] showed that the wind velocity has a direct correlation with the depth and speed of fading, and an inverse correlation with the quantity P_r . This area of work also includes the study described in article [83], whose articles obtain the relationship of P_r to meteorological conditions by using weather charts and data on the vertical velocity of air currents. A clear connection between the vertical velocity of air currents and the mean hourly values of P_r was discovered. Cloudiness has also been noted to have an effect on P_r [84].

II. THEORY

The three years from 1957 to 1959 represented a period in which the original form of the theories of incoherent scattering [53, 85-92, 106] was improved, and also a period in which new mechanisms and theories of long-range tropospheric propagation of ultrashort waves were investigated.

1. CONCEPT OF INCOHERENT SCATTERING. GLOBULAR INHOMOGENEITIES.

The reviews which have been especially compiled [93, 94, 52, 43, 12, 11, 9, 95, 96] or included in many works free us from the necessity of presenting the theory and basic principles of the concept of incoherent scattering. Because of the naturalness of this concept and its fundamental correspondence to the changeable nature of the "transdiffraction" field, this concept became extremely popular by the beginning of the 1950's. Thus, in the Proc. IRE (1955, No. 10) only a single article [97] was based on another mechanism of propagation. During this period, the layer structure of the troposphere was either totally ignored, or declared to be ineffective or random. This explains, for instance, why the quantity $(\Delta\epsilon)^2$ (or $(\Delta\epsilon)^2/l$) was made to vary according to a power of the height h^{-n} , although this scarcely agreed with experiments. Indeed, as a result of this choice, it was possible to obtain the much stronger correlation between P_r and distance required by experiments, and the theoretical curves could be adapted to the experimental curves by varying N .

The divergence of different theories with regard to the question of the nature of the variation of P_r/P_0 with frequency and the obscurity surrounding this question (which is still unclear, see above) also made it necessary to conduct further theoretical research. As soon as the turbulent nature of atmospheric fluctuations was generally recognized, considerable attention began to be devoted to problems in the theory of and in particular to the search for new correlation functions. The use of the transformation of a correlation function into a Fourier integral or series (spectral representation) provided a rigorous and convenient tool

of examination, presenting considerable heuristic value. For instance, the author of [118] arrives at the conclusion that the existence of a definite relationship between λ and the spectral component l is important in propagation problems, although the practical value of this idea is still unclear.

During the past three years, two well-known theories have been chiefly used in works devoted to questions in incoherent scattering: on the one hand, the Kolmogorov-Obukhov theory (see also [99, 100]), and on the other, the theory of "scrambling of the gradient" introduced by Villars and Weisskopf. The first theory uses a factor of $\lambda^{1/3}$ to express the frequency dependence of P_r/P_0 , while the second theory uses λ . Since the latter type of variation is in closer agreement with experiments, many people prefer the second theory. In particular, on the basis of this theory a hypothesis regarding the influence of the stratosphere in propagation over distances greater than 600 - 800 km has been proposed.

So far there has been no general solution of the Maxwell equations for a case of a statistically inhomogeneous medium above a spherical Earth. This general solution would obviously have to combine the theory of refraction propagation above the spherical Earth with the reemission in the fluctuating medium. So far the studies conducted in this direction hold only for special cases [102, 103] and for a plane Earth.

Because of the still insurmountable mathematical difficulties all theories of incoherent scattering tend first to radically simplify the problem so that it can be handled by the methods of geometric optics (the intersection of sharply limited radiation patterns in the refracting volume of the troposphere, i.e., they assume that all the intermediate regions do not play an important role in propagation).

It is known that, in effect, all such theories lead to equations of the "radar formula type"

$$\frac{P_r}{P_0} = Q D^2 \int_V \frac{\sigma(\theta)}{R_1^2 R_2^2} dV, \quad (1)$$

where Q is a constant factor; $\sigma(\theta)$ — "the scattering cross-section" — is some function expressing in one way or another the influence of the fluctuation of ϵ and the dependence on λ and the gradient $d\epsilon/dh$ (if only by reducing the radius of the earth to an effective radius). From this formal point of view the different theories differ basically only in the form of the function σ . The differences in the method of carrying out the integration in (1) and in the assumptions regarding the size and shape of the re-radiating volume V are far less essential than the assumptions regarding the variation of $(\Delta\epsilon)^2$ or $(\Delta\epsilon)^2/l$ with height or the quantitative value of these quantities.

Following Waterman [104], we can confine ourselves to the remarks that in all such theories $\sigma(\theta)$ is expressed essentially in the following simple way:

$$\sigma(\theta) = \frac{b}{\sin^m\left(\frac{\theta}{2}\right)},$$

where θ is the angle of reradiation, which can be taken equal to the angular distance between the receiver and transmitter; and b is an expression accounting for the dependence on $(\Delta\epsilon)^2$, l , $d\epsilon/dh$, etc. For even integers $m > 2$, a general formula is obtained in a good approximation in [104] by integrating Eq. (1) over all space above the planes which are tangential to the earth at the transmitter and the receiver:

$$\frac{P_r}{P_0} = Q b A_m D^{-m+3}, \quad (2)$$

where A_m depends on m . If $b \sim h^{-n}$, D^{-m+3} must be replaced by $D^{-m+3-2n}$. For theories in which m , while not an integer, is still close to an integral number, it is possible to replace m by the nearest integer (as for instance in [58]). The existing theories give results extremely close to Eq (2), and the difference between these results scarcely has any significance beyond a purely theoretical interest.

A characteristic of all the theories is the variation of the field as a power of D . Since often this does not correspond to experiment, it is not only necessary to introduce the

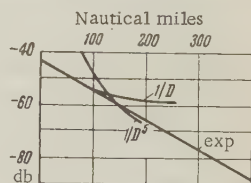


Figure 9. Comparison of different analytic functions expressing the field in terms of distance.

proportionality $b \sim h^{-n}$, but also to take different values of m for different distances. For instance, in [58] it was necessary to use the wholly artificial hypothesis, which does not correspond to the fact, that up to $h = 660$ m the quantity b is constant and is proportional to h^{-2} . However, (Figure 9) in the diagram ($\log P_r, D$) the points of the experimental curve fit easily onto an exponential curve, and if the field is taken to be an exponential function of D , the need for an artificial hypothesis regarding the variation with altitude disappears. It is true of a large number of experimental curves (see above).

For a comparison with experiment, it is also necessary to select the numerical value of $(\overline{\Delta\epsilon})^2$ or $(\overline{\Delta\epsilon})^2/l$ (which, it is true, lies within the limits of the observed values in the majority of cases). In this connection, there are no less than four parameters in [104] which can be selected arbitrarily.

2. CONCEPT OF INCOHERENT SCATTERING. TURBULENT LAYERS.

The structure of the troposphere in layers and the reflection (even though the special type) of waves from the troposphere's layers has long been proposed as a possible mechanism of long-range propagation. For instance, Feinstein's hypothesis [105] that the waves are reflected from an inhomogeneous atmosphere varying monotonically with height can be included in this category. When direct measurements showed that fairly stable tropospheric layers of the inhomogeneities did exist [71], the interest in the layer structure of the troposphere greatly increased.

One of the first to direct attention to the possible role of the layers in long-range propagation of ultrashort waves was Troitskiy [107] in 1955, who proposed a layer arrangement of the turbulent atmospheric inhomogeneities and obtained theoretical equations which corresponded to experiment (under certain assumptions). Schunemann [108] also developed the idea of the influence of layer-type inhomogeneities. The nature of the refraction from layers was developed by Brekhovskikh [109] and Beckmann [44]. Staras [60] who proposed considering the anisotropy of inhomogeneities can also be regarded as an initiator of this hypothesis to a certain extent. First, Friiff, Crawford, and Hogg [110] developed this idea clearly. In their treatment the chief role is played by the number and shape of the layer formation. They determined the reflection of waves by the methods of geometric optics combined with a determination of the coefficient of diffraction by the methods of Fresnel diffraction. The actual turbulence of the layers is not really taken into account, but it is assumed that individual layers can be oriented in any direction, so that layers lying outside the great circle of the duct may participate in the reflection to the receiver. The coefficient of reflection obtained is proportional to the change $[d\epsilon/dh]_0$ at the boundary of the layer. From a few assumptions partly arbitrary, about the numerical values of this change, the dimensions of the layers, their shapes and the shape of the volume V , a function which varies with distance according to θ^{-4} (i.e., D^{-4}) and which is proportional to λ is obtained. This function contains a considerable number of arbitrary parameters, providing great opportunities for bringing the theories into agreement with experiments. In a second very extensive work [5], a number of related questions is considered.

The concept of turbulent layers is being developed in detail by Castell, Misme, Voge, and Pizzichino [10, 111, 112, 113, 119] who have proposed a model of the atmosphere called a "petaled atmosphere." This model is also of the layer type, but consists of a large number of leaves, some of which are "quiet," and some of which are "disturbed." They support this hypothesis with meteorological observation and arrive at the following conclusion. The quiet layers provide chiefly mirror reflection, while the disturbed layers provide chiefly diffuse reflection. However, both mirror-reflection and diffuse-reflection components exist for each type of layer. The relative magnitude of these components is determined by the parameter

$$a = \frac{\sin \psi}{\lambda} \left[2\pi h + L \frac{\Psi + \psi}{2} \right],$$

where L is the "longitudinal size of the reflecting surface" (assumed to be finite), while $90^\circ - \Psi$ and $90^\circ - \varphi$ are the angles of slide and reflection ($\Psi = \varphi$ only in the case of mirror reflection). The smaller h and Ψ (and the larger λ), the larger and more constant is the mirror component. As a decreases, the straight line on the "Rayleigh diagram" indicating an exact Rayleigh distribution begins to bend somewhat and its slope becomes flatter and flatter, until finally it becomes a horizontal straight line, showing that the amplitude is constant [32

and [44]). For the "partial" field of individual layers, the mirror component of the field is found to decrease as the angle of slide Ψ and the frequency increase, while the diffuse component behaves in the reverse fashion. It is found that there is no need for a reflecting layer with a thickness h to be continuous and that the distinction between mirror reflection and diffuse reflection (and the corresponding field distributions) still holds for a "combination of reflectors" randomly distributed in the "h layer," even when the boundary surfaces are not flat. Mirror reflection disappears for $h < \frac{1}{2}\lambda$.

In a later work [119], this concept is developed. The fundamental significance of considering the instantaneous signal characteristics rather than the usually observed average values is emphasized. Formulas are derived for the power reflected from regular and irregular surfaces. These formulas are necessary for the subsequent computation of the received field. It is found that the three possible mechanisms of the propagation (mirror reflection, scattering from layers, and globular scattering) transform into one another as the frequency increases. This "provides a bridge" between the concept of partial reflection and scattering.

3. CONCEPT OF COHERENTLY REFLECTING LAYERS

A number of works postulate fairly long (ideally extending the entire length of the duct) and relatively stable layers of inhomogeneities of ϵ or the gradient* $d\epsilon/dh$, which scatter coherently the energy arriving from the transmitter.

These views, which have long been held with regard to the ionosphere, were first applied to the troposphere in the theory of tropospheric waveguides in the beginning of the 'forties. Then, following [105] Carroll and Ring developed these views. These authors [97], showed that, when the classical problem of diffraction is solved for a profile which is more complicated than a linear profile (leading to the concept of the equivalent radius) and taking into account [98], a very large number of terms in the series representing natural waves, which is a solution of the original differential equation, it is possible to obtain a result in satisfactory agreement with experiments. Ponomarenko [114] arrived at similar results by using Fok's diffraction theory [115]. Kalinin [66] applied the method of geometric optics to a troposphere with a monotonic profile (under certain assumptions) and obtained a positive result distinguished by very simple computational formulas.

The notion of the reflection of radio waves from stable layers in the troposphere encounters a double objection. The first objection concerns the incompatibility of the large variability of the field over large distances with the concept of stable tropospheric layers. However, by following a model such as that given in [112], variability can also be introduced into these layers without any special difficulty (although this can be done only theoretically for the present). The second objection is much more serious and applied to studies such as [97, 98]. However, this objection is rather of a theoretical nature and does not involve the practical aspect of the question very much. The point is that the authors of these studies assume the following type of function for ϵ : 1) decreases monotonically according to some equation up to $h = H_0$ (it is assumed that $H_0 = 9300$ m), and 2) at $h = H_0$ the quantity ϵ becomes constant (for instance, equal to unity) with a sharp discontinuity of the gradient $d\epsilon/dh$ at $h = H_0$. However, this means that at $h = H_0$ there is actually some (mathematically infinitely thin) layer of the gradient $d\epsilon/dh$. Indeed, Northover [116] arrived at the conclusion that there can be no distant field if the profile of $\epsilon(h)$ does not have this discontinuity. However, in a second study [117], this same author showed that when there is a layer $d\epsilon/dh$ at some height $h = H_1$ (H_1 not necessarily equal to H_0), it is possible to obtain a field of the necessary order at large distances. For sufficiently large D , the attenuation of P_r with distance D is stronger, the shorter λ . Generally speaking, this is in agreement with experiments. However, this result is attained at the price of introducing two arbitrary parameters: H_1 , and the intensity of the layer (i.e., the value of $d\epsilon/dh$ in the layer). In addition, both the actual computations and the final equations are much more complicated than in theories based on incoherent scattering.

Nonetheless, this theory deserves great attention. Indeed the actual existence of layers is reliably established by direct measurement (see above), while their role in long-range

*As well as of higher derivatives: $d^2\epsilon/dh^2$, etc.

propagation is established by studies of the dependence of P_r on distance. Thus, the authors [15] arrive at a definite positive conclusion regarding the existence and influence of such layers in their experiments at frequencies of 91.3 Mc. On their curve, there are clear maxima (observed by Megow [53] previously). These maxima are explained by the theory developed in [15], in which the methods of geometric optics are used for the reflection of radiowaves in stable spherical layers.

It is difficult to doubt that the diffraction approach (as distinguished from the method of geometric optics) to these problems is considerably more rigorous when applied to the problem of propagation along the curvature of the terrestrial sphere. Therefore, we made an attempt to simplify the equations given in [116, 117]. By the approximate method of replacing the summation of the infinite series in [116] and [117] by integration, an equation of the type

$$\frac{P_r}{P_0} = \frac{1}{D} \Phi \left(\lambda, \left[\frac{de}{dh} \right]_0, h_1, h_2 \right) \exp \{ -\alpha D \},$$

is obtained, where Φ is a complicated function similar to the height factors in the classical theory of diffraction which expresses the dependence on λ and on the magnitude of the discontinuity $[d\epsilon/dh]_0$ in the layer, while α is another function of the form $A - B \ln \lambda$ which depends on the same parameters (see above), where the quantities A and B do not depend on λ .

Without being rigorous and without showing explicitly the fluctuating nature of the field, this equation does, however, give the required exponential change of $P_r(D)$. (Since the variation with D^{-1} is unimportant relative to the exponential factor).

It is interesting to note, since $\alpha \sim \ln \lambda$, the variation with frequency is basically of the form $P_r/P_0 \sim \lambda^{BD}$.

It turns out that $BD \approx 1$ for $D = 300$ km. This is in agreement with the results of article [38].

CONCLUSION

The practical utilization of the phenomenon of long-range propagation is making large steps forward. However, for a complete theoretical understanding of the question, there remains much to do, in particular in regard to the study of the structure of the troposphere, the determination of the possible role of the stratosphere, the better understanding of the influence of climatic conditions, and so on.

The following questions have been little studied so far: the dependence of the signals on distance and frequency; the width of the passband; and the role of dispersing the antennas in space and with respect to frequency and angle of incidence. From the theoretical point of view, this is the result of the insufficient knowledge of the interaction of the three known mechanisms of propagation. Attention must be directed to this problem as well as to the possible extension of various old theories, at the same time as broad experiments are intensified and experience is acquired in operation.

In this way it will be possible to obtain the goal of the theoretical studies — reliable engineering formulas.

REFERENCES

1. J. H. Chisholm, *Onde electr.*, 1957, 37, 362, 427.
2. K. Bullington, *Proc. I. R. E.*, 1955, 43, 19, 1175.
3. G. Millington, *Electr. Engng*, 1958, 30, 363, 248.
4. R. A. Bowden, L. F. Taholm, J. W. Stark, *Proc., I. E. E.*, 1958, B 105, Suppl. No. 8, 84.
5. A. B. Crawford, D. C. Hogg, W. H. Kummer, *Bell System Techn. J.*, 1959, 38, 5, 1067.
6. I. Grosskopf, *Nachrichtentechnik*, 1956, 9, 6.
7. B. A. Vvedenskiy, and A. V. Sokolov, *Radiotekhnika i Elektronika*, 1957, 3, 11, 1375.
8. A. I. Kalinin, *Elektrosvyaz*, 1956, 5, 37.
9. D. M. Vysokovskiy, *Some Problems in the Long-Range Tropospheric Propagation of Ultrashort Waves*. Izd, AN SSSR, 1958.

10. F. du Castel, P. Misme, J. Voge, *Onde electr.*, 1958, 38, 376, 503.
11. A. G. Arenberg, *The Propagation of Decimeter and Centimeter Waves*.
12. B. A. Vvedenskiy, A. G. Arenberg, *Radiotekhnika*, 1957, 12, 1, 3, 10.
13. K. S. Jowet, *Proc. I. E. E.*, 1958 B 105, Suppl. No. 8, 91.
14. F. A. Kitchen, B. C. Richards, I. J. Richmond, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 108.
15. R. I. Starkey, W. R. Tarner, S. R. Racoe, G. F. Kitchen, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 97.
16. W. R. R. Joy, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 158.
17. W. R. R. Joy, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 143.
18. G. G. Rider, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 143.
19. B. C. Angell, I. B. L. Foot, W. J. Lucas, G. T. Thompson, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 128.
20. A. T. Waterman, *IRE Trans.*, 1958, AP-6, 4, 338.
21. H. E. Dinger, W. E. Gardner, D. H. Hamilton, A. E. Teachman, *Proc. I. R. E.*, 1958, 46, 7, 1401.
22. I. H. Vogelmann, I. L. Ryerson, M. H. Bickelhaupt, *Proc. I. R. E.*, 1959, 47, 5, 688.
23. L. A. Ames, E. T. Martin, E. J. Rogers, *Proc. I. R. E.*, 1959, 47, 5, 769.
24. H. Beverage, E. Laport, L. Simpson, *RCA Rev.*, 1955, 16, 3, 432.
25. Radio transmission by ionospheric and tropospheric scatter, *Proc. I. R. E.*, 1960, 48, 1, 30.
26. L. H. Doherty, G. A. Neal, *IRE Trans.*, 1959, AP-7, 2, 117.
27. B. Josephson, G. Carlson, *IRE Trans.*, 1958, AP-6, 2, 173.
28. L. Klinker, *UKW Fernempfangsbeobachtungen, ihre Bedeutung für Meteorologie und Funktechnik*, Berlin, 1955.
29. K. Bullington, W. J. Inkster, A. L. Durkee, *Proc. I. R. E.*, 1955, 43, 10, 1306.
30. S. O. Rice, *Proc. I. R. E.*, 1953, 41, 2, 274.
31. K. Norton, P. L. Rice, L. E. Vogler, *Proc. I. R. E.*, 1955, 43, 10, 1488.
32. K. Norton, L. E. Vogler, W. V. Mansfield, P. J. Short, *Proc. I. R. E.*, 1955, 43, 10, 1354.
33. G. C. Rider, *Marconi Rev.*, 1958, 21, 131, 184.
34. Discussion on tropospheric propagation beyond the horizon, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 184.
35. I. H. Gerks, *Proc. I. R. E.*, 1955, 43, 10, 1290.
36. K. Bullington, *Bell System Techn. J.*, 1957, 36, 3, 593.
37. *Proc. of the XII General Assembly URSL*, 1957, 11, part 2.
38. R. Bolgiano, *Proc. I. R. E.*, 1959, 47, 2, 331.
39. D. L. Ringwalt, W. S. Ament, T. G. Macdonald, *IRE, Trans.*, 1958, AP-6, 2, 208.
40. W. Morrow, *Onde electr.*, 1957, 37, 362, 444.
41. M. W. Gough, *Marconi Rev.*, 1958, 21, 131, 198.
42. J. H. Chisholm, P. A. Portmann, J. T. Bettencourt, J. T. Roche, *Proc. I. R. E.*, 1955, 43, 1317.
43. J. B. Wilsner, A. J. Pote, *Onde electr.*, 1957, 37, 362, 456.
44. P. Beckmann, *Acta Technica CSAV*, 1957, 2, 4, 311.
45. L. G. Trolese, *Proc. I. R. E.*, 1955, 43, 10, 1300.
46. J. H. Chicholm, L. P. Rainville, J. F. Roche, H. G. Root, *IRRE Trans.*, 1959, CS-7, 3, 195.
47. I. Kurihawa, *Proc. I. R. E.*, 1955, 43, 10, 1362.
48. G. L. Mellen, W. E. Morrow, A. J. Pote, W. Radford, J. B. Wiesner, *Proc. I. R. E.*, 1955, 43, 10, 1269.
49. W. M. Sharpless, *Proc. I. R. E.*, 1946, 34, 11, 837.
50. A. B. Crawford, W. M. Sharpless, *Proc. I. R. E.*, 1946, 34, 10, 845.
51. I. A. Gusyatinskiy, B. S. Nadenenko, *Elektrosvyaz*, 1959, 12, 10.
52. F. du Castel, *Onde electr.*, 1957, 37, 362, 591.
53. E. C. S. Megow, *Nature*, 1950, 166, 12, 1100.
54. P. L. Rice, A. G. Longley, K. A. Norton, *NBC Report*, 1958, 5582, 30.
55. N. A. Armand, B. A. Vvedinskiy, *Radiotekhnika i Elektronika*, 1959, 4, 10, 1594.
56. F. du Castel, *Onde electr.*, 1960, 40 394, 10.
57. R. Cabessa, *Onde electr.*, 1960, 40, 394 24.
58. E. G. Richards, *Proc. I. E. E.*, 1958, B 105, Suppl. No. 8, 177.
59. L. Boithias, F. du Castel, *Onde electr.*, 1960, 40, 394, 39.
60. H. Staras, *Proc. I. R. E.*, 1955, 43, 10, 1374.
61. J. H. Chisholm, L. P. Rainville, J. F. Roche, H. G. Root, *IRE Trans.*, 1958, AP-6, 4, 377.

62. W. H. Tidd, Proc. I. R. E., 1955, 43, 10, 1297.
63. H. G. Booker, J. T. de Bettencourt, Proc. I. R. E., 1955, 43, 10, 281.
64. W. E. Gordon, Proc. I. R. E., 1955, 43, 110, 23.
65. V. N. Troitskiy, Radiotekhnika, 1956, 11, 9, 3.
66. A. J. Kalinin, Elektrosvyaz, 1959, 6, 41.
67. I. A. Gusyatinskiy, Elektrosvyaz, 1959, 4, 3.
68. A. B. Prosin, V. F. Gubskiy, Radiotekhnika, 1959, 14, 5, 23.
69. J. Voge, Ann. telecommun., 1953, No. 8-9.
70. L. A. Kazakov, A. N. Lomakin, Proceedings of the Conference on Radio Engineering and Communication, Gopov Institute, 1959.
71. C. Crain, Proc. I. R. E., 1955, 43, 10, 1405.
72. C. Crain, Onde electr., 1957, 37, 362, 441.
73. L. Klinker, Z. Meteorol., 1957, 11, 2, 8, 43.
74. L. Klinker, Z. Meteorol., 1957, 11, 10, 8, 339.
75. B. Bean, B. Cahoon, Proc. I. R. E., 1957, 45, 11, 1545.
76. B. H. Bean, G. D. Thayer, Proc. I. R. E., 1959, 47, 5, 740.
77. B. Bean, IRE Trans., 1956, CS-4, 1, 32.
78. B. Bean, J. D. Horn, J. Res. Nat. Bur. of Standards, 1959, 63D, 3, 259.
79. J. Bonaveglia, Alta Frequenza, 1958, 27, 6, 815.
80. P. Misme, IRE Trans., 1958, AP-6, 3, 289.
81. L. Klinker, Hochfrequenztechn. und Elektroakust., 1956, 65, 3, 77.
82. D. R. Hay, G. E. Poaps, Canad. J. Phys., 1959, 37, 11, 1272.
83. W. Moler, W. Arvola, Trans. Amer. Geophys. Union, 1956, 399.
84. P. Beckmann, Slaboproudy obzor, 1955, 16, 3, 115.
85. V. A. Krasil'nikov, Izvestiya Akad. Nauk SSSR, Ser. geogr. i geofiz., 1949, 13, 3, 33.
86. C. L. Pekeris, Phys. Rev., 1947, 71, 4, 268.
87. H. G. Booker, W. E. Gordon, 1950, 38, 4, 401.
88. H. Staras, J. Appl. Phys. 1952, 43, 10, 1152.
89. R. A. Silverman, M. Balzer, Phys. Rev., 1954, 96, 11, 560.
90. A. D. Wheelon, Proc. I. R. E., 1955, 43, 10, 1381.
91. A. V. Prosin, Proceedings of the Conference of Radio Engineering and Communication, Gopov Institute, 1959, 3.
92. G. S. Gorelik., Radiotekhnika i Elektronika, 1956, 17, 4, 695.
93. P. Beckmann, Slaboproudy obzor, 1956, 17, 4, 188.
94. H. Staras, A. D. Wheelon, Proc. of the XII General Assembly URSI, 1957, 11, p. 2, com. II, 85.
95. V. J. Tatarski, Theory of Fluctuation Phenomena during Wave Propagation in Turbulent Atmosphere, 1959.
96. N. G. Denisov, and V. A. Zverev, Izvestiya Vuzov MVO SSSR (Radiofizika), 1959, 2, 4, 521.
97. T. J. Carroll, R. M. Ring, Proc. I. R. E., 1955, 43, 10, 1384.
98. T. J. Carroll, R. M. Ring, Onde electr., 1957, 37, 362.
99. R. A. Silverman, Proc. I. R. E., 1955, 43, 10, 1253.
100. R. A. Silverman, J. Appl. Phys. 1956, 27, 7, 669.
101. H. G. Booker, W. E. Gordon, Proc. I. R. E., 1957, 45, 9, 1233.
102. N. G. Denisov, Izvestiya Vuzov MVO SSSR (Radiofizika), 1958, 1, 5-6, 41.
103. F. G. Bass, and E. A. Kanner, Izvestiya Vuzov MVO SSSR (Radiofizika), 1959, 2, 553.
104. A. T. Waterman, Proc. I. R. E., 1958, 46, 11, 1842.
105. J. Feinstein, J. Appl. Phys., 1951, 22, 10, 1292.
106. J. A. Saxston, Proc. I. E. E., 1951, 98, 5, 360.
107. V. N. Troitskiy, Radiotekhnika, 1956, 11, 5, 3.
108. R. Schunemann, Hochfrequenztechn. und Elektroakust., 1957, 66, 2, 52; 1958, 67, 6.
109. A. M. Brekhovskikh, Waves in stratified media. Published by the Acadmeny of Sciences of the USSR, 1957.
110. H. T. Friis, A. B. Crowford, D. C. Hogg, Bell System Techn. J., 1957, 36, 627.
111. F. du Castel, P. Misme, J. Voge, Ann. telecommun., 1958, 13, 7-8, 207.
112. F. du Castel, P. Misme, J. Voge, Ann. telecommun., 1958, 13, 9-10, 265.
113. F. du Castel, P. Misme, J. Voge, Ann. telecommun., 1959, 14, 1-2, 35.
114. L. M. Ponomarenko, Radiotekhnika i elektronika, 1959, 4, 6, 930.
115. V. A. Fok, Izvestiya Akad. Nauk SSSR, Ser. Fiz., 1959, 14, 1, 70.
116. F. H. Northover, Canad. J. Phys., 1955, 33, 5, 241.

117. F. H. Northover, *Canad. J. Phys.*, 1955, 33, 6, 316, 347.
118. E. D. Denman, *Proc. I. R. E.*, 1960, 48, 1, 112.
119. F. du Castel, P. Misme, A. Pizzichino, I. Voge, *Ann. telecomuns*, 1960, 15, 1-2, 38.

Received July 27, 1960

PROPAGATION OF ELECTROMAGNETIC WAVES OVER A GYROTROPIC MEDIUM

N. D. Khaskind

This article deals with a plane surface of separation between a vacuum and a homogeneous absorbing anisotropic medium. An electromagnetic field, excited by given sources, is studied in the vacuum above this anisotropic medium.

Generalized boundary conditions are used for the solution of this problem. On this basis, the propagation of electromagnetic waves formed by elementary electric and magnetic radiators is analyzed. The greatest attention is devoted to transversely magnetized plasma and to ferrite. For these cases, the fields are determined as RMS values. This makes it extremely simple to find the reflection equations and the attenuation functions for different polarizations. The results obtained have particular significance for the analysis of the propagation of radio waves near the gyrotropic ionosphere.

1. BASIC EQUATIONS

Suppose we have a homogeneous absorbing anisotropic medium, for which we write the complex dielectric and magnetic tensors in the general form

$$\hat{\epsilon} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix}, \quad \hat{\mu} = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_{33} \end{pmatrix} \quad (1)$$

In the future, we shall account for the time dependence by means of the exponential factor $\exp i\omega t$, which we have omitted for brevity. We shall use the rationalized practical system of units, in which ϵ_0 and μ_0 represent the permittivity and permeability, respectively, while the primed symbols ϵ'_{ij} and μ'_{ij} refer to relative permittivity and relative permeability respectively. We direct the z -axis along the outward normal to the boundary of the anisotropic medium, and we locate the xy -plane on this boundary.

An electromagnetic field penetrates only a small depth into an absorbing medium. Therefore, for the tangential components of the intensity of the electromagnetic field at the surface of separation, the same equation can be adopted as for plane waves propagating in an anisotropic medium along the normal to this surface of separation. The specific form of these

equations is easily established, although it is first necessary to carry out a series of tedious computations with the general equations for an electromagnetic field in an anisotropic medium for the case of plane waves propagating in the negative direction of the z -axis. We shall omit these computations, and proceed from the following approximate boundary conditions. For the determination of the electromagnetic field in a vacuum above an absorbing anisotropic medium*:

$$E_x = -\rho_0 (\rho_{11} H_x + \rho_{12} H_y), \quad E_y = -\rho_0 (\rho_{21} H_x + \rho_{22} H_y) \quad \text{for } z = 0 \quad (2)$$

($\rho_0 = (\mu_0 / \varepsilon_0)^{1/2} = 120\pi$ ohms), where the dimensionless coefficients ρ_{pq} are defined as follows:

$$\begin{aligned} \rho_{11} &= \frac{\alpha_{11}(\mu) n_1 n_2 - \alpha_{22}(\varepsilon) \beta(\mu)}{n_1 n_2 (n_1 + n_2)}, & \rho_{12} &= \frac{\alpha_{12}(\mu) n_1 n_2 - \alpha_{12}(\varepsilon) \beta(\mu)}{n_1 n_2 (n_1 + n_2)}, \\ \rho_{21} &= -\frac{\alpha_{21}(\mu) n_1 n_2 - \alpha_{21}(\varepsilon) \beta(\mu)}{n_1 n_2 (n_1 + n_2)}, & \rho_{22} &= -\frac{\alpha_{22}(\mu) n_1 n_2 - \alpha_{11}(\varepsilon) \beta(\mu)}{n_1 n_2 (n_1 + n_2)}. \end{aligned} \quad (3)$$

Here n_1 and n_2 are the complex refractive indexes of the ordinary and extraordinary waves:

$$\begin{aligned} n^4 - \gamma n^2 + \beta(\varepsilon) \beta(\mu) &= 0, & n_{1,2}^2 &= \frac{1}{2} (\gamma \pm (\gamma^2 - 4\beta(\varepsilon) \beta(\mu))^{1/2}) \\ (n_{1,2} &= n_{1,2r} - i n_{1,2i}, & n_{1,2i} &> 0). \end{aligned} \quad (4)$$

For the sake of abbreviation, we shall adopt the following notation ($a = \varepsilon$ or μ):

$$\begin{aligned} \alpha_{11}(a) &= a'_{21} - \frac{a_{31}}{a_{33}} a'_{23}, & \alpha_{12}(a) &= a'_{22} - \frac{a_{32}}{a_{33}} a'_{23}, & \alpha_{21}(a) &= a_{11} - \frac{a_{31}}{a_{33}} a_{13}, \\ \alpha_{22}(a) &= a'_{12} - \frac{a_{32}}{a_{33}} a'_{13}, & \beta(a) &= \alpha_{11}(a) \alpha_{22}(a) - \alpha_{12}(a) \alpha_{21}(a), \\ \gamma &= \alpha_{12}(\mu) \alpha_{21}(\varepsilon) + \alpha_{31}(\mu) \alpha_{12}(\varepsilon) - \alpha_{11}(\mu) \alpha_{11}(\varepsilon) - \alpha_{22}(\mu) \alpha_{22}(\varepsilon). \end{aligned} \quad (5)$$

The coefficient ρ_{pq} can be represented in a somewhat different form. For instance, because of the equation $n_1^2 n_2^2 = \beta(\varepsilon) \beta(\mu)$, we have for ρ_{11}

$$\rho_{11} = \frac{\alpha_{11}(\mu) \beta(\varepsilon) - \alpha_{22}(\varepsilon) n_1 n_2}{\beta(\varepsilon) (n_1 + n_2)}. \quad (6)$$

In the special case of anisotropic absorbing medium, we obtain

$$n_1 = n_2 = (\varepsilon' \mu')^{1/2}, \quad \rho_{11} = \rho_{22} = 0, \quad \rho_{12} = -\rho_{21} = \left(\frac{\mu'}{\varepsilon'} \right)^{1/2}. \quad (7)$$

Equations (2) then reduce to Leontovich's boundary condition. In the same way as Leontovich's boundary conditions, boundary condition (2) are correct for refractive indexes whose modulus is large relative to unity ($|n_{1,2}| \gg 1$). However, an exact determination of the order of the neglected powers of the quantities is more difficult here than in the case of an isotropic medium.

The approximate boundary conditions (2) are useful in that they describe immediately the characteristic properties of the refraction of plane waves, namely linearly polarized plane waves falling on an anisotropic surface acquire a transverse elliptic polarization in reflection. Indeed, suppose that E-polarized waves fall on an anisotropic surface (Fig. 1);

$$E_n = e^{-ik(x \sin \theta + z \cos \theta)}, \quad (8)$$

* The general form of these conditions is given in [1].

where k is the wave number, and θ is the angle of incidence (for convenience we have directed the z -axis into the anisotropic medium and therefore it is necessary to reverse the sign in the right-hand side of condition (2)). We obviously have two mutually perpendicular transverse components for the refracted wave

$$E^r = R_e^{(1)} e^{-ik(x \sin \theta - z \cos \theta)}, \quad E_y^r = R_e^{(2)} e^{-ik(x \sin \theta - z \cos \theta)}, \quad (9)$$

where $R_e^{(1)}$ and $R_e^{(2)}$ are the coefficients of reflection of the electric field. By using the well-known equations relating E and H in a plane wave, as well as conditions (2), we easily establish

$$R_e^{(1)} = \frac{(\cos \theta - \rho_{12})(1 - \rho_{21} \cos \theta) - \rho_{11} \rho_{22} \cos \theta}{(\rho_{12} + \cos \theta)(1 - \rho_{21} \cos \theta) + \rho_{11} \rho_{22} \cos \theta},$$

$$R_e^{(2)} = \frac{2\rho_{22} \cos \theta}{(\rho_{12} + \cos \theta)(1 - \rho_{21} \cos \theta) + \rho_{11} \rho_{22} \cos \theta}. \quad (10)$$

In exactly the same way, when H-polarized waves fall on the surface (2)

$$E_{ny} = e^{-ik(x \sin \theta + z \cos \theta)}, \quad E_y^r = R_m^{(1)} e^{-ik(x \sin \theta - z \cos \theta)}, \quad (11)$$

$$E^r = R_m^{(2)} e^{-ik(x \sin \theta - z \cos \theta)}$$

we obtain similar equations for the coefficients of reflection

$$R_m^{(1)} = \frac{\rho_{11} \rho_{22} \cos \theta - (1 + \rho_{21} \cos \theta)(\rho_{12} + \cos \theta)}{(\rho_{12} + \cos \theta)(1 - \rho_{21} \cos \theta) + \rho_{11} \rho_{22} \cos \theta},$$

$$R_m^{(2)} = \frac{2\rho_{11} \cos \theta}{(\rho_{12} + \cos \theta)(1 - \rho_{21} \cos \theta) + \rho_{11} \rho_{22} \cos \theta}. \quad (12)$$

This explains, in particular, at least from the qualitative point of view, the property of the propagation of long radio waves, which acquire an elliptic polarization in reflection from the ionosphere, as is well known practice.

Let us obtain the values of the coefficients ρ_{pq} for a gyrotropic electron plasma. For transverse magnetization of the plasma, the dielectric tensor $\hat{\epsilon}'_z$ becomes [2]

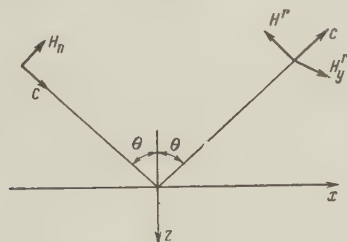


Fig. 2.

$$\hat{\epsilon}'_z = \begin{pmatrix} \xi & i\eta & 0 \\ -i\eta & \xi & 0 \\ 0 & 0 & \epsilon'_e \end{pmatrix}, \quad (13)$$

$$\xi = 1 + \frac{q^2(1-is)}{\sigma^2 - (1-is)^2}, \quad \eta = \frac{\sigma q^2}{\sigma^2 - (1-is)^2}$$

$$\epsilon'_e = 1 - \frac{q^2}{1-is}, \quad s = \frac{\nu}{\omega}, \quad \sigma = \frac{\omega_H}{\omega},$$

$$q = \frac{\omega_p}{\omega}, \quad \omega_p^2 = \frac{Ne^2}{m_e \epsilon_0}, \quad \omega_H = \frac{eB_0}{m_e}.$$

Here e and m_e are the electron charge and mass; $B_0 = \mu_0 H_0$ is the induction of the magnetizing field; ν is the number of collisions of the electron with heavy particles per unit time; ω_p is the plasma frequency; ω_H is the gyromagnetic frequency; and N is the concentration of free electrons. Taking $\mu'_{pq} = \delta_{pq}$ for the plasma, where δ_{pq} is the unit tensor, and using Eq. (3) - (5), we obtain

$$\rho_{11} = \rho_{22} = \frac{i}{2} \left(\frac{1}{n_2} - \frac{1}{n_1} \right) = g, \quad \rho_{12} = -\rho_{21} = \frac{1}{2} \left(\frac{1}{n_2} + \frac{1}{n_1} \right) = r, \quad (14)$$

$$n_1^2 = \xi + \eta, \quad n_2^2 = \xi - \eta.$$

The coefficients of reflection become

$$R_e^{(1)} = \frac{\cos^2 \theta - \left(r - \frac{1}{r} + \frac{g^2}{r} \right) \cos \theta - 1}{\cos^2 \theta + \left(r + \frac{1}{r} + \frac{g^2}{r} \right) \cos \theta + 1},$$

$$R_m^{(1)} = \frac{\cos^2 \theta + \left(r - \frac{1}{r} + \frac{g^2}{r} \right) \cos \theta - 1}{\cos^2 \theta + \left(r + \frac{1}{r} + \frac{g^2}{r} \right) \cos \theta + 1}, \quad (15)$$

$$R_e^{(2)} = R_m^{(2)} = \frac{2 \frac{g}{r} \cos \theta}{\cos^2 \theta + \left(r + \frac{1}{r} + \frac{g^2}{r} \right) \cos \theta + 1}.$$

Suppose now that we have a plasma magnetized longitudinally in the direction of the x-axis:

$$\hat{\epsilon}'_x = \begin{pmatrix} \epsilon'_e & 0 & 0 \\ 0 & \xi & i\eta \\ 0 & -i\eta & \xi \end{pmatrix}. \quad (16)$$

Then from Equations (3) - (5), (10) and (12), we find

$$\rho_{11} = \rho_{22} = 0, \quad \rho_{12} = \frac{1}{n_2} = \rho_1, \quad \rho_{21} = -\frac{1}{n_1} = -\rho_2, \quad n_1^2 = \frac{\xi^2 - \eta^2}{\xi}, \quad n_2^2 = \epsilon'_e, \quad (17)$$

$$R_e^{(1)} = \frac{\cos \theta - \rho_1}{\cos \theta + \rho_1}, \quad R_m^{(1)} = \frac{\rho_2 \cos \theta - 1}{\rho_2 \cos \theta + 1}, \quad R_e^{(2)} = R_m^{(2)} = 0. \quad (18)$$

With regard to the ionospheric layer D, it is necessary to take into account the isotropic role played by the ions in this layer. As a constant, ξ is replaced by $\xi_0 = \xi + \epsilon_i$ in transverse magnetization, where ϵ_i is the complex relative dielectric constant caused by the ionized part of the D-layer. In longitudinal magnetization, we have

$$n_1^2 = \xi - \frac{\eta^2}{\xi} + \epsilon'_i, \quad n_2^2 = \epsilon'_e + \epsilon'_i.$$

Similar equations hold for gyrotropic ferrite. Indeed, in longitudinal magnetization of ferrite, the magnetic tensor μ_z is of the same form as ϵ_z , i.e.,

$$\hat{\mu}'_z = \begin{pmatrix} \xi_1 & i\eta_1 & 0 \\ -i\eta_1 & \xi_1 & 0 \\ 0 & 0 & \mu' \end{pmatrix}, \quad (19)$$

for which the values of ξ_1 and η_1 are given in [3] with an allowance for the absorption. From Eqs. (3) - (5), we obtain

$$\rho_{11} = \rho_{22} = -\frac{i}{2\epsilon'} (n_1 - n_2) = g, \quad \rho_{12} = -\rho_{21} = \frac{1}{2\epsilon'} (n_1 + n_2) = r, \quad (20)$$

$$n_1^2 = \epsilon' (\xi_1 + \eta_1), \quad n_2^2 = \epsilon' (\xi_1 - \eta_1).$$

For longitudinal magnetization of the ferrite along the x-axis, we have

$$\rho_{11} = \rho_{22} = 0, \quad \rho_{12} = \frac{n_1}{\epsilon'} = \rho_1, \quad \rho_{21} = -\frac{n_2}{\epsilon'} = -\rho_2,$$

$$n_1^2 = \frac{\epsilon'}{\xi} (\xi_1^2 - \eta_1^2), \quad n_2^2 = \epsilon' \mu'. \quad (21)$$

2. FIELD OF ELEMENTARY RADIATORS

To determine the excited electromagnetic field above a gyrotropic surface, we shall use Hertz electric and magnetic vectors Π_e and Π_m :

$$\mathbf{E} = \text{grad div } \Pi_e - k^2 \Pi_e - ik\rho_0 \text{rot } \Pi_m, \quad (22)$$

$$\mathbf{H} = \text{grad div } \Pi_m - k^2 \Pi_m - \frac{ik}{\rho_0} \text{rot } \Pi_e.$$

Let us suppose that the radiators are directed along the x- and z-axes ($\Pi_e = \Pi_{ex} = 0$), and let us consider the cases of transverse and longitudinal magnetization separately. In the first case, in view of (14) and (20), boundary conditions (2) become

$$E_x = -\rho_0 (gH_x + rH_y), \quad E_y = \rho_0 (rH_x - gH_y) \quad \text{for } z = 0. \quad (23)$$

As a result of the substitution of (22) into (23), the following equations are obtained for $z = 0$:

$$M_1 \Pi_{ez} = -\frac{\partial \Pi_{ex}}{\partial x} - ik\rho_0 s_0 \Pi_{mz}, \quad M_2 \Pi_{mz} = -\frac{\partial \Pi_{mx}}{\partial x} - \frac{ik}{\rho_0} s_0 \Pi_{ez}, \quad (24)$$

$$M_2 \Pi_{ex} = ik\rho_0 s_0 \Pi_{mx}, \quad M_1 \Pi_{mx} = \frac{ik}{\rho_0} s_0 \Pi_{ex} \quad (s_0 = \frac{g}{r}), \quad (25)$$

where M_1 and M_2 are differential operators

$$M_1 = \frac{\partial}{\partial z} - ik \left(r + \frac{g^2}{r} \right), \quad M_2 = \frac{\partial}{\partial z} - \frac{ik}{r}. \quad (26)$$

In the second case, on the basis of (17) and (21), conditions (2) are outwardly more simple:

$$E_x = -\rho_0 \rho_1 H_y, \quad E_y = \rho_0 \rho_2 H_x \quad \text{for } z = 0. \quad (27)$$

However, after substituting (22) into (27), we obtain the following equations for $z = 0$:

$$\frac{\partial}{\partial x} \left(\frac{\partial \Pi_{ex}}{\partial x} + \frac{\partial \Pi_{ez}}{\partial z} - ik\rho_1 \Pi_{ez} \right) = -\rho_0 \rho_1 \left(\frac{\partial \Pi_{mx}}{\partial x} + \frac{\partial \Pi_{mz}}{\partial z} - \frac{ik}{\rho_1} \Pi_{mz} \right), \quad (28)$$

$$\frac{\partial}{\partial y} \left(\frac{\partial \Pi_{ex}}{\partial x} + \frac{\partial \Pi_{ez}}{\partial z} - ik\rho_2 \Pi_{ez} \right) = \rho_0 \rho_2 \left(\frac{\partial \Pi_{mx}}{\partial x} + \frac{\partial \Pi_{mz}}{\partial z} - \frac{ik}{\rho_2} \Pi_{mz} \right), \quad (29)$$

$$\frac{\partial \Pi_{ex}}{\partial z} - \frac{ik}{\rho_1} \Pi_{ex} = 0, \quad \frac{\partial \Pi_{mx}}{\partial z} - ik\rho_2 \Pi_{mx} = 0. \quad (30)$$

Despite the fact that boundary conditions (27) are outwardly more simple than boundary conditions (23), the determination of the field for the case of longitudinal magnetization is nonetheless more complicated than the determination of the field for transverse magnetization. In our subsequent discussion, we shall consider only normal (transverse) magnetization. For longitudinal magnetization or for the general case, it is necessary to use standard methods based on Sommerfeld's formula for the expansion of a spherical wave and on subsequent analysis of the multiple integrals.

Let us consider first a transverse electric radiator, setting

$$\Pi_{ez}^{(1)} = \frac{e^{-ikr_1}}{r_1} - \frac{e^{-ikr_2}}{r_2} = \varphi_e,$$

$$\Pi_{mz}^{(1)} \neq 0, \quad \Pi_{ex}^{(1)} = \Pi_{mx}^{(1)} = 0, \quad (31)$$

$$r_1 = (r_0^2 + (z - z_0)^2)^{1/2}, \quad r_2 = (r_0^2 + (z + z_0)^2)^{1/2}, \quad r_0^2 = x^2 + y^2,$$

where r_1 and r_2 are the distances between the point B(x, y, z) and the points A(0, 0, z_0) and A*(0, 0, $-z_0$) (Fig. 3), while the functions φ_e and $\Pi_{mz}^{(1)}$ satisfy the wave equation and are regular for all $z \geq 0$.

From conditions (24), we obtain the equation

$$M_1 \varphi_e = -ik\rho_0 s_0 \Pi_{nz}^{(1)} + 2 \frac{\partial}{\partial z} \left(\frac{e^{-ikr_2}}{r_2} \right), \quad M_2 \Pi_{nz}^{(1)} = -\frac{ik}{\rho_0} s_0 \varphi_e, \quad (32)$$

which are correct throughout all space they are regular. Accordingly after eliminating the function φ_e , we find

$$\begin{aligned} M \Pi_{nz}^{(1)} &= -\frac{2}{\rho_0} ik s_0 \frac{\partial}{\partial z} \left(\frac{e^{-ikr_2}}{r_2} \right), \\ M &= M_1 M_2 - (ik s_0)^2 = \\ &= \left(\frac{\partial}{\partial z} - ik p_1 \right) \left(\frac{\partial}{\partial z} - ik p_2 \right), \end{aligned} \quad (33)$$

where p_1 and p_2 are the roots of the characteristic equation

$$p^2 - \left(r + \frac{1}{r} + \frac{g^2}{r} \right) p + 1 = 0; \quad (34)$$

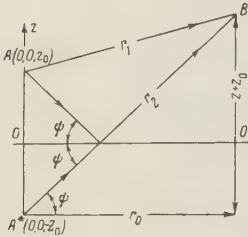


Fig. 3.

$$p_{1,2} = \frac{1}{2} \left[r + \frac{1}{r} + \frac{g^2}{r} \pm \left(\left(r + \frac{1}{r} + \frac{g^2}{r} \right)^2 - 4 \right)^{1/2} \right].$$

Because of the absorption, p_1 and p_2 are always complex quantities. As shown in [4], (see also [2], [5]), Equation (33) must be extended analytically to the complex plane $z = z_r + iz_i$ to avoid the appearance of free solutions. The complete solution to this equation which will satisfy the wave equation can now be determined from the contour integrals

$$\Pi_{nz}^{(1)} = -\frac{2s_0 ik}{\rho_0 (p_1 - p_2)} \left(p_1 e^{ikp_1 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_1 \zeta + r'_2)}}{r'_2} d\zeta - p_2 e^{ikp_2 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_2 \zeta + r'_2)}}{r'_2} d\zeta \right) \quad (35)$$

($r'_2 = (r_0^2 + (\zeta + z_0)^2)^{1/2}$), in which the lower limit of integration is equal to $i\infty$ for $\text{Re } p_{1,2} < 0$, and to $i\infty$ for $\text{Re } p_{1,2} > 0$.

Let us now use second equation of (32) and also Eq. (31). We find

$$\begin{aligned} \Pi_{ez}^{(1)} &= \frac{e^{-ikr_1}}{r_1} + \frac{e^{-ikr_2}}{r_2} + \frac{2}{p_1 - p_2} \left(ik p_1 \left(p_1 - \frac{1}{r} \right) e^{ikp_1 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_1 \zeta + r'_2)}}{r'_2} d\zeta - \right. \\ &\quad \left. - ik p_2 \left(p_2 - \frac{1}{r} \right) e^{ikp_2 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_2 \zeta + r'_2)}}{r'_2} d\zeta \right). \end{aligned} \quad (36)$$

Suppose now that we have a transverse magnetic radiator, for which we assume

$$\begin{aligned} \Pi_{ez}^{(2)} &\neq 0, \quad \Pi_{nz}^{(2)} = \frac{e^{-ikr_1}}{r_1} - \frac{e^{-ikr_2}}{r_2} + \varphi_m, \\ \Pi_{ex}^{(2)} &= \Pi_{mx}^{(2)} = 0, \end{aligned} \quad (37)$$

where $\Pi_{ez}^{(2)}$ and φ_m are regular functions satisfying the wave equation for all $z \geq 0$. Conditions (24) lead to the equation

$$\begin{aligned} M \Pi_{ez}^{(2)} &= -2ik\rho_0 s_0 \frac{\partial}{\partial z} \left(\frac{e^{-ikr_2}}{r_2} \right) \\ M_1 \Pi_{ez}^{(2)} &= -ik\rho_0 s_0 \varphi_m, \end{aligned} \quad (38)$$

from which we find

$$\begin{aligned} \Pi_{ez}^{(2)} &= \rho_0^2 \Pi_{nz}^{(1)}, \quad \Pi_{nz}^{(2)} = \frac{e^{-ikr_1}}{r_1} + \frac{e^{-ikr_2}}{r_2} + \\ &\quad \frac{2ik}{p_1 - p_2} \left(p_1 \left(p_1 - r - \frac{g^2}{r} \right) e^{ikp_1 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_1 \zeta + r'_2)}}{r'_2} d\zeta - \right. \\ &\quad \left. - p_2 \left(p_2 - r - \frac{g^2}{r} \right) e^{ikp_2 z} \int_{\pm i\infty}^z \frac{e^{-ik(p_2 \zeta + r'_2)}}{r'_2} d\zeta \right). \end{aligned} \quad (39)$$

The fields of longitudinal electric and magnetic radiators are determined in exactly the same way. Thus we have the following equations:

$$\Pi_{ex}^{(3)} = \Pi_{mx}^{(2)}, \quad \Pi_{mx}^{(3)} = -\Pi_{mz}^{(1)}, \quad \Pi_{ez}^{(3)} = -\frac{\rho_0}{iks_0} \frac{\partial \Pi_{mx}^{(3)}}{\partial x}, \quad \Pi_{mz}^{(3)} = 0, \quad (40)$$

$$\Pi_{ex}^{(4)} = -\Pi_{ez}^{(2)}, \quad \Pi_{mx}^{(4)} = \Pi_{ez}^{(1)}, \quad \Pi_{ez}^{(4)} = 0, \quad \Pi_{mz}^{(4)} = -\frac{1}{ik\rho_0 s_0} \frac{\partial \Pi_{ex}^{(4)}}{\partial x}, \quad (41)$$

where $\Pi_e^{(3)}$, $\Pi_m^{(3)}$, and $\Pi_e^{(4)}$, $\Pi_m^{(4)}$ are the Hertz vectors corresponding to the electric and magnetic longitudinal radiators.

When these formulas are used, it should be remembered that r and g are small quantities for a plasma and that they determine the order of accuracy of boundary condition (23). At the same time, from (14) and (34), it is easily established that

$$p_{1,2} = \frac{1 + n_1 n_2 \left(1 \pm \left(\left(1 - \frac{1}{n_1^2} \right) \left(1 - \frac{1}{n_2^2} \right) \right)^{1/2} \right)}{n_1 + n_2}. \quad (42)$$

Hence, it is seen that p_1 is of the order of the quantities n_1 and n_2 , which are large in modulus, while p_2 is of the same order of smallness of the quantities r and g . Therefore to maintain the same accuracy in Eqs. (35), (36), (39), (40) and (41) as in boundary conditions (23), it is necessary to use the following expansion for large $|p_1|$:

$$-ikp_1 \int_{\pm i\infty}^z e^{-ikp_1(\zeta-z)} U(r_0, \zeta) d\zeta = U(r_0, z) + \frac{1}{ikp_1} \frac{\partial U}{\partial z} + \dots \quad (43)$$

3. REFLECTION EQUATIONS AND ATTENUATION FUNCTIONS.

The definite integrals appearing in (35), (36) and (39) for the Hertz vectors are of the same form as for the special case of propagation of radio waves above a semiconducting earth. Consequently the same approximate methods for the analysis of the propagation of radio waves [2] can be used to obtain refraction equation and attenuation functions in the given problem. For this purpose, let us consider the integral

$$F = ikpe^{ikpz} \int_{\pm i\infty}^z \frac{e^{-ik(p\zeta+r_2')}}{r_2'} d\zeta. \quad (44)$$

Forming the substitution of variables $\zeta - z = \tau$ in (44) we obtain

$$F = ikp \int_{\pm i\infty}^0 \frac{e^{-ik(p\tau+r_2')}}{r_2'} d\tau, \quad (45)$$

where we have for r_2' in accordance with Fig. 3

$$(r_2')^2 = r_2^2 + 2r_2\tau \sin \psi + \tau^2 \quad \left(\psi = \frac{\pi}{2} - \theta \right).$$

The expression under the integral in (45) decays exponentially as $\tau \rightarrow \pm i\infty$. Therefore, for an approximate computation of (45), we can say

$$r_2' \simeq r_2 + \tau \sin \psi + \frac{\tau^2}{2r_2} \cos^2 \psi,$$

After which we obtain for F

$$F \simeq ikp \frac{e^{-ikr_2}}{r_2} \int_{\pm i\infty}^0 e^{-ik((p+\sin\psi)\tau + \cos^2\psi(\tau^2/2r_2))} d\tau. \quad (46)$$

In the case of high corresponding points, we can take $\cos^2\psi \simeq 0$. Then, from (46) we find

$$F \simeq -\frac{p}{p + \sin\psi} \frac{e^{-ikr_2}}{r_2}. \quad (47)$$

This expression converts (35), (36), and (39) into the reflection equations.

$$\begin{aligned} \Pi_{ez}^{(1)} &= \frac{e^{-ikr_1}}{r_1} + R_e^{(1)} \frac{e^{-ikr_2}}{r_2}, \quad \Pi_{mz}^{(1)} = \frac{1}{\rho_0} R_e^{(2)} \frac{e^{-ikr_2}}{r_2}, \\ \Pi_{mz}^{(2)} &= \frac{e^{-ikr_1}}{r_1} + R_m^{(1)} \frac{e^{-ikr_2}}{r_2}, \quad \Pi_{ez}^{(2)} = \rho_0 R_m^{(2)} \frac{e^{-ikr_2}}{r_2}, \end{aligned} \quad (48)$$

where the coefficients of reflection are defined by Eq. (15).

Let us now examine another limiting case, when $\cos^2\psi = 1$, i.e., $z = z_0 = 0$, $\sin\psi = 0$, $r_1 = r_2 = r_0$. Setting

$$U = (-ik)^{1/2} \left(\frac{\tau}{(2r_0)^{1/2}} - \left(\frac{r_0}{2} \right)^{1/2} p \right),$$

we can transform (46) in the form

$$F = L(x) \frac{e^{-ikr_0}}{r_0}, \quad L(x) = -2x^{1/2} e^{-x} \int_{\pm i\infty}^{x^{1/2}} e^{U^2} dU, \quad x = -\frac{i}{2} p^2 k r_0. \quad (49)$$

Let us substitute (49) into (35), (36) and (39) (after dividing by 2). We then find

$$\begin{aligned} \Pi_{ez}^{(1)} &= \frac{e^{-ikr_0}}{r_0} V_e^{(1)}, \quad \Pi_{mz}^{(1)} = \frac{e^{-ikr_0}}{\rho_0 r_0} V_e^{(2)}, \\ \Pi_{mz}^{(2)} &= \frac{e^{-ikr_0}}{r_0} V_m^{(1)}, \quad \Pi_{ez}^{(2)} = \rho_0 \frac{e^{-ikr_0}}{r_0} V_m^{(2)}, \end{aligned} \quad (50)$$

where V_e and V_m are the attenuation functions, defined by the equation

$$\begin{aligned} V_e^{(1)} &= 1 + \frac{1}{p_1 - p_2} \left(\left(p_1 - \frac{1}{r} \right) L(x_1) - \left(p_2 - \frac{1}{r} \right) L(x_2) \right), \\ V_e^{(2)} &= V_m^{(2)} = -\frac{s_0}{p_1 - p_2} (L(x_1) - L(x_2)) \quad (x_{1,2} = x(p_{1,2})), \\ V_m^{(1)} &= 1 + \frac{1}{p_1 - p_2} \left(\left(p_1 - r - \frac{g^2}{r} \right) L(x_1) - \left(p_2 - r - \frac{g^2}{r} \right) L(x_2) \right). \end{aligned} \quad (51)$$

Here the dimensionless quantities $x_{1,2}$ can be regarded as "numerical" distances in the gyrotropic case.

If $|x_{1,2}| \gg 1$, it is easily shown that

$$-L(x) \sim 1 - \frac{1}{2x} - \frac{3}{4x^2} - \dots$$

by integrating by parts. In this case, we obtain the following asymptotic equations:

$$V_e^{(1)} = -\frac{1 - g^2}{x_0 r^2}, \quad V_e^{(2)} = V_m^{(2)} = -\frac{g(g^2 + r^2 + 1)}{x_0 r^2}, \quad (52)$$

$$V_m^{(1)} = \frac{1 - (g^2 + r^2)(g^2 + r^2 + 1)}{x_0 r^2} \quad (x_0 = -ikr_0),$$

These show that in the gyrotropic case the field intensity decreases as r_0^{-2} , i.e., in the same way as in an isotropic absorbing surface. Let us also note that in the general case expression (46) can be transformed into a form similar to (49).

REFERENCES

1. L. L. Landau, E. M. Lifshits, *Electrodynamics of Continuous Media* FIFML, 59, p. 397.
2. Ya. L. Al'pert, V. L. Ginzburg, E. L. Feynberg, *Propagation of Radio Waves*, GTTI, 1953, pp. 152-157, 326.
3. G. Sool, L. Walker, *Waveguide Propagation of Electromagnetic Waves in Gyrotropic Media*, IL, 1955, p. 13.
4. G. D. Malyuzhinets, On a generalization of Weyl's Formula for a wave field above an absorbing surface, *AN SSSR*, 1948, 60, 3, 367.
5. M. D. Khaskind, Excitation of surface EM waves on flat dielectric sheaths, *Radiotekhnika i Elektronika*, 1960, 5, 2, 188.

Received May 9, 1960

ASYNCHRONOUS OSCILLATIONS IN SELF-EXCITED OSCILLATOR WITH TWO DEGREES OF FREEDOM IN OVERVOLTAGE OPERATION

A. N. Bruyevich

The report discusses overvoltage operation in an asynchronous oscillator with two degrees of freedom. It is considered that overvoltage operation derives from the screen grid and does not affect the value of bias. Dependence of amplitude of oscillations on circuit parameters is discussed. It is noted that frequency jumps are possible with a change in resonance resistances of tuned circuits or in feedback factors.

It is shown that, due to overvoltage operation, there appear in the parameter planes new regions in which single frequency or double-frequency oscillation may occur, depending on the initial conditions.

INTRODUCTION

The processes occurring in asynchronous self-excited oscillators with two degrees of freedom upon the onset of overvoltage operation were not studied until the present time. The author has succeeded in conducting such investigation with approximation of the anode and grid currents in straight line segments. Based on reference [1], this investigation dealt with the forming of pulses of anode current of a tube in overvoltage operation from peaked pulses, which permits reducing the expression for the mean transconductance of the anode current in overvoltage operation to the algebraic sum of the mean transconductance of the peaked pulses as discussed in detail in [2]. The present work is a continuation of [2] and permits study of those regions of the parameter plane and the plane of amplitudes which [2] included in the overvoltage condition. In addition, a number of new phenomena are noted.

1. MEAN SLOPE OF OVERVOLTAGE OPERATION

The initial equations are the equations for slowly varying amplitudes and the equation for a circuit with lagless self-bias and have the form (notation is the same as that used in [2])

$$T_i \frac{dU_i}{dt} + U_i = R_i I_{2-i, i-1}, \quad i = 1, 2, \quad (1)$$

$$E_c = -I_{c00} R_c.$$

Using the method of double Fourier series, we shall examine the derivation of the frequency dependence of the first harmonic of the anode current of the i -th tuned circuit $I_{2-i, i-1}$ on amplitudes at the grid U_i and the bias voltage E_c .

The total voltage at the grid and anode of the tube may be represented in the form

$$e_c = E_c + U_1 \cos \omega_1 t + U_2 \cos \omega_2 t = E_c + U_1 \cos \tau_1 + U_2 \cos \tau_2, \quad (2)$$

$$e_a = E_a - U_{a1} \cos \tau_1 - U_{a2} \cos \tau_2.$$

Considering one of the voltages (e.g., at the second tuned circuit) to be constant, we obtain for the voltages at the grid and at the anode

$$e_c = E_{c1} + U_1 \cos \tau_1, \quad e_a = E_{a1} - U_{a1} \cos \tau_1, \quad (3)$$

as long as it is established that

$$E_{c1} = E_c + U_2 \cos \tau_2, \quad E_{a1} = E_a - U_{a2} \cos \tau_2. \quad (4)$$

Shaping of the anode pulse upon the action of the voltage of (3) at the grid and anode was discussed in detail in [1]. We will note only that this pulse has the form shown in Figure 1 (heavy line). It is not difficult to see from Figure 1 that the pulse may be "constructed" from four peaked pulses.

Accordingly, the analytical expression for amplitude of the n -th harmonic of plate current has the form

$$I_n = S(U_1 - DU_{a1})\gamma_n(\theta_0) - S(U_1 - DU_{a1})\gamma_n(\theta_1) - S_{\kappa}U_{a1}\gamma_n(\theta_1) + S_{\kappa}U_{a1}\gamma_n(\theta_2). \quad (5)$$

Here, as in reference [1], S is the transconductance, S_{κ} is the transconductance of the line of critical operation, $\gamma_n(\theta)$ is the coefficient of expansion of the n -th harmonic for a peaked pulse.

If, in accordance with [1], we designate $S_{\kappa}/S = v + D$ and recall that for a self-excited oscillator with grid and anode voltages are related by the coefficient of transformation k ($k = U_{a1}/U_1$, the reciprocal of the feedback factor), then (5) may be presented in the form (for the first harmonic)

$$I_1 = SU_1 [(1 - Dk)\gamma_1(\theta_0) - (1 + vk)\gamma_1(\theta_1) + vk\gamma_1(\theta_2)]. \quad (6)$$

It was shown in [1] that the expressions for the cosines of the cutoff angles have the form

$$\cos \theta_0 = -\frac{E_{c1} - E'_c}{U_1 - DU_{a1}}, \quad \cos \theta_1 = -\frac{E_{c1} - E'_c - (D + v)E_{a1}}{U_1 + vU_{a1}}, \quad (7)$$

$$\cos \theta_2 = \frac{E_{a1}}{U_{a1}}.$$

Now let us recall that the grid and anode voltages of the tube vary with the frequency of the second tuned circuit (4). Herein all three cutoff angles become variable in time and the cosines of these angles behave as instantaneous dimensionless voltages acting on the expansion coefficients of the peaked pulses of anode current (for each of the pulses these coefficients serve as modulation characteristics). From (4), expressions (7) for the instantaneous values of the cosines of the cutoff angle may be given in the form

$$\cos \theta_i = y_i - x_i \cos \tau_2, \quad i = 0, 1, 2, \quad (8)$$

where

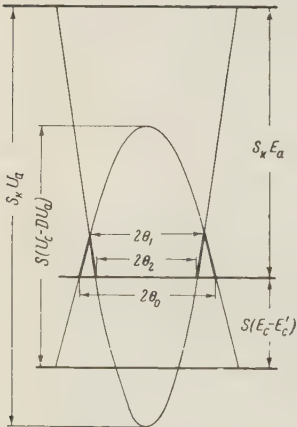


Figure 1. Shaping of anode current pulse in overvoltage operation.

$$\begin{aligned}
y_0 &= -\frac{E_c - E'_c}{U_1(1 - Dk_1)}, & x_0 &= \frac{U_2(1 - Dk_2)}{U_1(1 - Dk_1)}, \\
y_1 &= -\frac{E_c - E'_c + (D + \nu)E_a}{U_1(1 + \nu k_1)}, & x_1 &= \frac{U_2(1 + \nu k_2)}{U_1(1 + \nu k_1)}, \\
y_2 &= \frac{\nu E_a}{\nu k_1 U_1}, & x_2 &= \frac{U_2 \nu k_2}{U_1 \nu k_1}.
\end{aligned} \tag{9}$$

Expressions y_i and x_i in (8) are dimensionless biases and the amplitude of the voltages acting on the modulation characteristics of peaked pulses.

Now, in order to obtain I_{10} (harmonics of anode current with frequency of the first tuned circuit) expression (6) must be averaged over time, taking into account that, in accordance with (8), the cutoff angles vary with the frequency of the second tuned circuit.

Introducing the mean transconductance S_{11} equal to the ratio of the amplitude of the first harmonic to the amplitude of the corresponding oscillation, we obtain

$$S_{11} = \frac{I_{10}}{S U_1 (1 - Dk_1)} = \overline{\gamma_1(\theta_0)} - \frac{1 + \nu k}{1 - Dk_1} \overline{\gamma_1(\theta_1)} + \frac{(D + \nu)k}{1 - Dk_1} \overline{\gamma_1(\theta_2)}. \tag{10}$$

The line over the expansion coefficient signifies averaging over time τ_2 . The first subscript in the expression for mean transconductance S_{11} denotes that this is the mean transconductance for the frequency of the first tuned circuit and the second subscript indicates that the case under discussion is that in which $U_1 > U_2$. If it is necessary to plot the mean transconductance for the case where $U_2 > U_1$ (i.e., S_{12}), then, as in [2], we first average the plate current pulse for the frequency of the second tuned circuit (that is, in (2) we first obtain the voltage $U_1 \cos \tau_1$) and then we isolate oscillations at the frequency of the first tuned circuit. We then obtain

$$S_{12} = \frac{I_{10}}{S U_1 (1 - Dk_1)} = \frac{1}{x_0} \widetilde{\gamma_0(\theta_0)} - \frac{1 + \nu k_1}{1 - Dk_1} \frac{1}{x_1} \widetilde{\gamma_0(\theta_1)} + \frac{(D + \nu)k_1}{1 - Dk_1} \frac{1}{x_2} \widetilde{\gamma_0(\theta_2)}, \tag{11}$$

where the wavy line over the expansion coefficient indicates isolation of the first harmonic for the frequency of the first tuned circuit. The expressions for the instantaneous values of the cosine of the cutoff angle have the same form as in (8) and we obtain the values of dimensionless bias y_i and amplitude x_i from (9) by changing the positions of subscripts 1 and 2.

As already stated, $\gamma_n(\theta_i)$ are the expansion coefficients for the peaked pulse and averaging and isolation of the first harmonic for this expansion coefficient were performed in [2] in investigating undervoltage operation. It was shown in [2] that

$$\overline{\gamma_1(\theta_i)} = S_{11}(x_i, y_i), \quad \widetilde{\gamma_0(\theta_i)} = S_{12}(x_i, y_i), \tag{12}$$

where $S_{11}(x_i, y_i)$ and $S_{12}(x_i, y_i)$ are the dimensionless mean transconductances of undervoltage operation. Hence, from (10) and (11), with consideration of (12), we obtain

$$S_{11} = S_{11}(x_0, y_0) - \frac{1 + \nu k_1}{1 - Dk_1} S_{11}(x_1, y_1) + \frac{(D + \nu)k_1}{1 - Dk_1} S_{11}(x_2, y_2), \tag{13}$$

$$S_{12} = S_{12}(x_0, y_0) - \frac{1 + \nu k_1}{1 - Dk_1} S_{12}(x_1, y_1) + \frac{(D + \nu)k_1}{1 - Dk_1} S_{12}(x_2, y_2). \tag{14}$$

Thus, the mean transconductance of overvoltage operation is a linear combination of the mean transconductances of undervoltage operation from different arguments. This is due to the fact that the current pulse in overvoltage operation was constructed as the algebraic sum of the peaked pulses.

Since the boundaries of overvoltage operation (signifying, also, the boundaries of applicability of formulas (13) and (14)) depend on the type of bias, we shall determine them for the case of automatic grid bias.

2. OVERVOLTAGE OPERATION WITH GRID BIAS

Let us plot mean transconductance as a function of bias for the case of overvoltage operation. We begin by noting that the second of the equations in (1), describing the state of the

self-bias circuit, remains the same as in [2] (the screen grid circuit is the overvoltage condition); hence, its solution (dependence of bias $y_c = -E_c/U_{1,2}$ on the amplitude ratio $x = U_{2,1}/U_{1,2}$ and on the bias factor $R_c S_c$) remains the same as in [2]. This solution, representing the detector characteristic of the grid circuit, is shown in Figure 2. It may be seen that the bias increases relative to one of the amplitudes $y_1 = E_1/U_{1,2}$ both with an increase in the second amplitude $x = U_{2,1}/U_{1,2}$ and with an increase in bias factor $R_c S_c$, which has obvious significance. With $R_c S_c = \infty$ we obtain a peak detector.

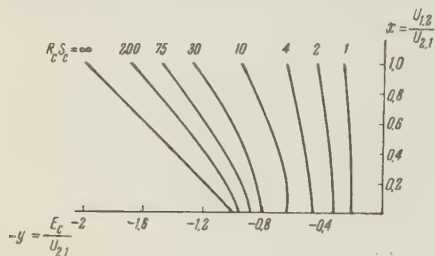


Figure 2. Detector characteristic of grid circuit.

$$\kappa_1 = \frac{1 - Dk_1}{1 - Dk_2} \frac{1 + vk_2}{1 + vk_1}, \quad \kappa_2 = \frac{1 - Dk_1}{1 - Dk_2} \frac{vk_2}{vk_1}, \quad \alpha_1 = \frac{1 - Dk_1}{1 + vk_1} \frac{(v + D) E_a}{-E'_c}, \quad \alpha_2 = \frac{1 - Dk_1}{(v + D) k_1} \frac{(v + D) E_a}{-E'_c}, \quad (16)$$

$$\beta_1 = \frac{1 - Dk_1}{1 + vk_1} \left[\frac{(v + D) E_a}{-E'_c} - 1 \right], \quad \beta_2 = \frac{1 - Dk_1}{(v + D) k_1} \frac{(v + D) E_a}{-E'_c},$$

and

$$y_c = -\frac{E_c}{U_{1,2}}.$$

For a pentode we may assume $D=0$, which considerably simplifies equality (16). In this case all six parameters may be expressed in terms of three, namely, vk_1 , vk_2 and $vE_a/-E'_c$.

If, in addition, we assume that bias y_c depends on the ratio of amplitudes x_0 and bias factor $R_c S_c$, then it develops that the mean transconductance of overvoltage operation is a function of two variables and four parameters (we omit zero in y_0 and x_0) $S(vk_1, vk_2, vE_a/-E'_c, R_c S_c, y, x)$.

Let us turn to an evaluation of the boundaries of overvoltage operation. We note that in the case of self-bias, for an increase in intensity of operation it is necessary to increase the geometric bias $(-E_c + E_c)$. This is because an increase in bias in this case (with constant $R_c S_c$) is possible only by increasing the voltage amplitude at the grid (meaning also the anode), which increases the intensity of operation. Thus, we must find the boundary value of bias y_{bd} at which for all $y > y_{bd}$ overvoltage operation will occur. Let us first find the boundary of low-overvoltage operation y_{1bd} (two cutoff angles) and then the boundary of high-overvoltage operation y_{2bd} (three cutoff angles). For this purpose let us examine Figure 1, from which it is evident that overvoltage operation occurs at the moment when tangency of the sinusoid occurs. It is obvious that the maximum probability of tangency exists at that moment when all four voltages have peak amplitude values, that is, when $\cos \tau_1 = \cos \tau_2 = 1$. With this in mind, we obtain from (8) for the boundary of low-overvoltage operation $y_1 - x_1 = 1$ or, from (15)

$$y_{1bd} = \frac{\alpha_1 y_c (R_c S_c, x) - \kappa_1 x - 1}{\beta_1}. \quad (17)$$

In precisely the same manner, for the boundary of high overvoltage operation we obtain

$$y_{2bd} = \frac{\alpha_2 y_c (R_c S_c, x) - \kappa_2 x - 1}{\beta_2}. \quad (18)$$

Thus, calculating from (16) the values of the circuit parameters, we may for each given value of x find $y_1 \text{ bd}$ and $y_2 \text{ bd}$. For $y < y_1 \text{ bd}$ operation will be undervoltage; for $y_1 \text{ bd} < y < y_2 \text{ bd}$ operation will be low overvoltage ($\theta_2 = 0$), and in order to plot the mean transconductance it is sufficient to use the first two terms of (13) or (14); for $y > y_2 \text{ bd}$ operation is high overvoltage and it is necessary to use all three terms.

It is still necessary to indicate the limits of applicability of such a plot. For this purpose we again refer to Figure 1. Since the voltages at the anode and grid of the tube vary with the frequency of the second tuned circuit, we shall attempt to follow these changes. For example, let the anode voltage E_a be reduced and the grid bias E_c be increased; this may be represented as that moment at which the cutoff angle θ_1 first becomes equal to and then greater than angle θ_0 . In an actual circuit at this moment the anode current will be zero, whereas our method of plotting the plate current pulse yields no pulse equal to zero. Hence, for those values of bias at which θ_1 may be greater than θ_0 our plot will not be accurate. Mathematically it is required that $\theta_0 \geq \theta_1$, that is, $\cos \theta_0 < \cos \theta_1$ or, with consideration of (8),

$$y_0 - x_0 \cos \tau_2 \leq y_1 - x_1 \cos \tau_2.$$

Inserting (15), we obtain for the critical value

$$y_{1 \text{ cr}} = \frac{\alpha_1 y_c (R_c S_c, x) - x |1 - \kappa_1|}{1 + \beta}, \quad (19)$$

where $|1 - \kappa_1|$ is the modulus of the expression $1 - \kappa_1$.

If $y > y_{1 \text{ cr}}$, at certain moments in time it will happen that $\theta_1 > \theta_0$. In this case the mean transconductance of overvoltage operation cannot be represented in the form of the sum of the mean transconductances of peaked pulses, for it is necessary to introduce new terms considering that our plot does not give for these moments a zero anode current.

If we further increase the bias (meaning, also, the amplitudes), it is then conceivable to have a condition where at all times $\theta_1 > \theta_0$. This means that the tube will be cut-off at all times and the mean transconductance will be equal to zero. For this purpose it is sufficient that $\cos \theta_0 > \cos \theta_1$, that is,

$$y_0 - x_0 \cos \tau_2 \geq y_1 - x_1 \cos \tau_2$$

or

$$y_{2 \text{ cr}} = \frac{\alpha_1 y_c (R_c S_c, x) + x |1 - \kappa_1|}{1 + \beta}. \quad (20)$$

Thus our plot is inaccurate for $y_{1 \text{ cr}} < y < y_{2 \text{ cr}}$; however, for identical coupling coefficients $k_1 = k_2$ both these values coincide: $y_{1 \text{ cr}} = y_{2 \text{ cr}}$, since in this case $\kappa_1 = 1$, that is, our method of plotting is accurate for any changes in y . Hence, for reasons of simplicity, we shall examine the case $k_1 = k_2$.

We may now plot the mean transconductance of overvoltage operation. This is most simply done in the following manner. For the given circuit let us find all parameters not depending on time as given in expression (16). We assign a value of $R_c S_c$, for the given ratio of amplitudes (x); we find the value of grid bias (y_c) (Figure 2), and from formulas (17) and (18) we determine the boundaries of overvoltage operation. We then assign a value of geometric bias (y) and by means of (15) obtain x_i, y_i ($i = 1, 2$). From the plots of mean transconductance of peaked pulse [2] we select the values of S_{11} (x_i, y_i). Combining these values in accordance with (13), we find the value of S_{11} for overvoltage operation. In the same manner we determine S_{12} (using the plots for S_{12} (x_i, y_i)). Mean transconductances plotted in this manner are shown in Figures 3a and 3b. The following parameters were adopted as the basis of calculation: $E_a = 400 \text{ v}$, $E_{c2} = 150 \text{ v}$, $E_c = -14 \text{ v}$, $S = 9 \text{ ma/v}$, $S_k = 2.3 \text{ ma/v}$, $R_c S_c = 50$, $k_1 = k_2 = 5$.

Let us examine, for example, the curve for $x = 1$ in Figure 3a. For values of $y < 0.95$ overvoltage operation occurs and the curve is the same as that given in [2]. For $0.95 < y < 1.1$ operation is low overvoltage and the curve falls rapidly (downward inflection); for $y > 1.1$ there appears a second cutoff angle and the curve begins to drop more slowly. The shape of the curves with other values of parameter x is the same.

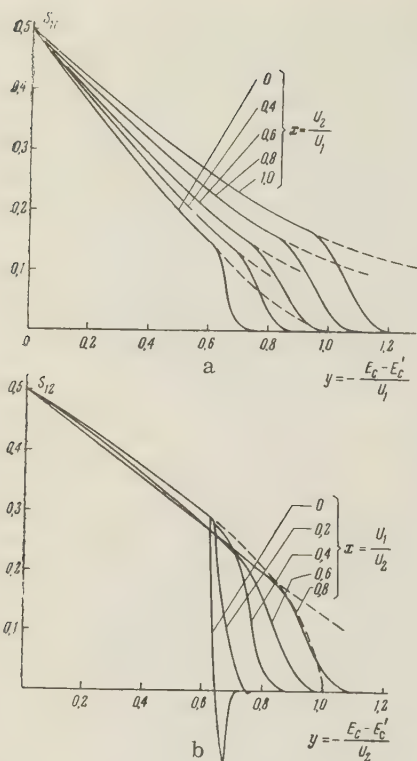


Figure 3. Mean transconductance as a function of bias voltage: a) large amplitudes in first tuned circuit; b) large amplitudes in second tuned circuit; dashed line, curves neglecting overvoltage operation.

For curves corresponding to the second form of entry (Figure 3b) it must be noted that with small value of x (when the amplitude of oscillations in the second tuned circuit is much greater than the amplitude of oscillations in the first tuned circuit) negative values of mean transconductance may be obtained (for example, curve $x = 0$ in Figure 3b).

For explanation of this, let us make use of the fact that, with $U_1 = 0$,

$$S_{12} = \frac{I_{10}}{U_1} = \frac{1}{U_1} \frac{2}{\pi} \int_0^{\pi} I_0(E_c + U_1 \cos \tau, U_2) \cos \tau d\tau \xrightarrow{U_1 \rightarrow 0} \frac{\partial I_0}{\partial U_1},$$

where I_0 is the direct component of current for frequency of the second tuned circuit (for proof we note that with $U_1 \rightarrow 0$ we obtain the indeterminate form $0/0$, which we evaluate by L'Hospital's rule and arrive at the above statement). This indicates that the mean transconductance depends on whether the direct component of current for frequency of the second tuned circuit increases or decreases with an increase in amplitude of oscillations in the first tuned circuit. Referring to Figure 1 (now E_a and E_c vary with the frequency of the first tuned circuit), we see that with an increase in amplitude of oscillations of the first tuned circuit angles θ_0 , θ_1 , and θ_2 , increases; hence, it is possible that the area delimited by the anode current pulse will increase or decrease. Since the zero component is proportional to this area, the mean transconductance may be positive or negative. This is quite clearly seen from Figure 3b (curve $x = 0$).

Thus, for $y < 0.65$ $S_{12} > 0$ and for $y > 0.65$ $S_{12} < 0$, which confirms the above remarks.

3. INVESTIGATION OF STABLE STATES

Assuming in stable operation that $dU_i/dt = 0$, we obtain from (1) for the steady-state solution equations for the principal isoclines at the plane of amplitudes (the second subscript for the mean transconductance is omitted since it is unimportant which amplitude is greater):

$$U_1 \left[S_1(x, y) - \frac{1}{R_1 S} \right] = 0, \quad (21)$$

$$U_2 \left[S_2(x, y) - \frac{1}{R_2 S} \right] = 0. \quad (22)$$

Let us plot the solution of (21) and (22) by the same method used in [2]. We will recall that this solution is to be sought as the intersection of curves (21) and (22). Plotting of each curve (e.g., for (21)) is performed as follows: the amplitude ratio (x) is given and from it for a given $R_1 S$ we find the geometric bias y , which is the abscissa of the point of intersection of curve S_{11} or S_{12} (in Figures 3a and 3b) and the straight line extended at the $1/R S$ level. Then the bias y_c for the given $R_c S_c$ is found from Figure 2. After this we find the amplitudes from the formulas

$$\frac{U_{1,2}}{-E'_c} = \frac{1}{y_c - y}, \quad \frac{U_{2,1}}{-E'_c} = x \frac{U_{1,2}}{-E'_c}. \quad (23)$$

Such a plot is given in Figure 4 (boundaries of the low- and high-overvoltage states are indicated by dashed line). The curves for overvoltage operation are compressed compared

with the corresponding curves for undervoltage operation and the curves extending to infinity vanish. Let us also note the following important fact. In undervoltage operation (for $R_1 S > 2$) the quench amplitude is greater than the amplitude of free oscillation (curve for $R_1 S = 3$ in Figure 4). In overvoltage operation, for sufficiently large $R_1 S$ ($R_1 S > 8$ in Figure 4) the amplitude of free oscillations is greater than the quench amplitude. This fact must be emphasized since it is of great importance in investigating the stability of double-frequency oscillations.

The curves in Figure 4 describe the behavior of a self-excited oscillator in the presence of a special form of external signal acting simultaneously on the anode and grid but out of phase by a ratio proportional to the coefficient of transformation.

Let us examine in detail one of the curves in Figure 4, for example, the curve for $R_1 S = 9$. With an increase in U_2 there is a decrease in U_1 . At first this is a gradual decrease, but subsequently it occurs more rapidly and for $U_2 / -E'_c > 4.5$ there occurs abrupt extinction of oscillation. We note also that for those curves where the function $U_1(U_2)$ is no longer single-valued (e.g., $R_1 S = 20$) the upper portions of the curves correspond to the stable state. Thus, for $R_1 S = 20$ the portion $U_1 / -E'_c > 1$. If U_2 exceeds a certain boundary value (for $R_1 S = 20$ $U_2 / -E'_c > 5$), interruption of oscillation occurs and at the moment of interruption the amplitude has a finite value. With a decrease in U_2 oscillation does not occur until U_2 reaches the value $U_2 / -E'_c > 4.75$, after which oscillation immediately occurs with large amplitude. Thus, in a self-excited oscillator with one degree of freedom, in the presence of external asynchronous action, hysteresis phenomena will be observed.

Now let us plot the solutions of Eq. (22), which may be obtained from (21) by changing the coordinate axes. This is possible in our case because the feedback factor is identical for both frequencies. The intersection of the curves plotted for Eqs. (21) and (22) gives us the points of stable operation. The possible types of such points are shown in Figure 5a-5g. The arrows in this figure indicate the field of directions of the integral curves of equation (1), from which we are able to evaluate the stability.

By varying the regeneration of one of the tuned circuits (for example, of the first $R_1 S$), with unchanged regeneration of the second tuned circuit ($R_2 S$), it is possible to proceed from the change in position of the singular points (that is, we may plot the dependence of oscillatory amplitudes on regeneration of the tuned circuits). Such dependence is plotted in Figure 6a-6e, where $\log R_1 S$ is given on the abscissa axis. Each plot in Figure 6 represents the dependence of oscillatory amplitude on regeneration in a self-excited oscillator with one degree of freedom (curve of free oscillation).

Since Figures 5 and 6 are thus related, we shall examine them together.

1. $R_2 S = 3$. The nature of the change in amplitude of oscillation with a change in regeneration of the first tuned circuit $R_1 S$ is seen from Figure 6a. This change may be observed also from Figure 5a by displacing the dashed curve in accordance with the change in $R_1 S$.

Figure 5a shows several dashed curves, which, with an increase in $R_1 S$ expand, as it were, from the coordinate origin.

The stable singular points are nodes (indicated in Figure 5 by Y) and the unstable points are saddles (C). At small values of $R_1 S$ (until the solid curve embraces the dashed curve) instead of saddle C_1 there will be a node Y_1 , that is, oscillations will occur in the second tuned circuit. With an increase in $R_1 S$ oscillations will occur in the first tuned circuit; their amplitude will increase and the amplitude of oscillations in the second tuned circuit will decrease (points Y'_1, Y''_1, Y'''_1). At large values of $R_1 S$ oscillations will exist only in the first tuned circuit (point Y'''_1). This is typical of the undervoltage condition.

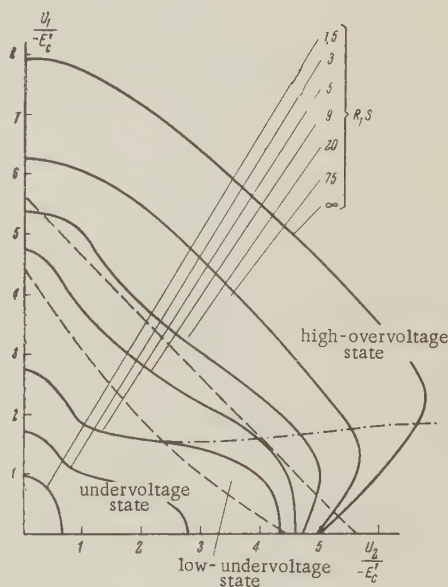


Figure 4. Principal isoclines of amplitude planes:

dashed lines, boundaries of low- and high-overvoltage operation; dash-dot line, curves disregarding overvoltage operation.

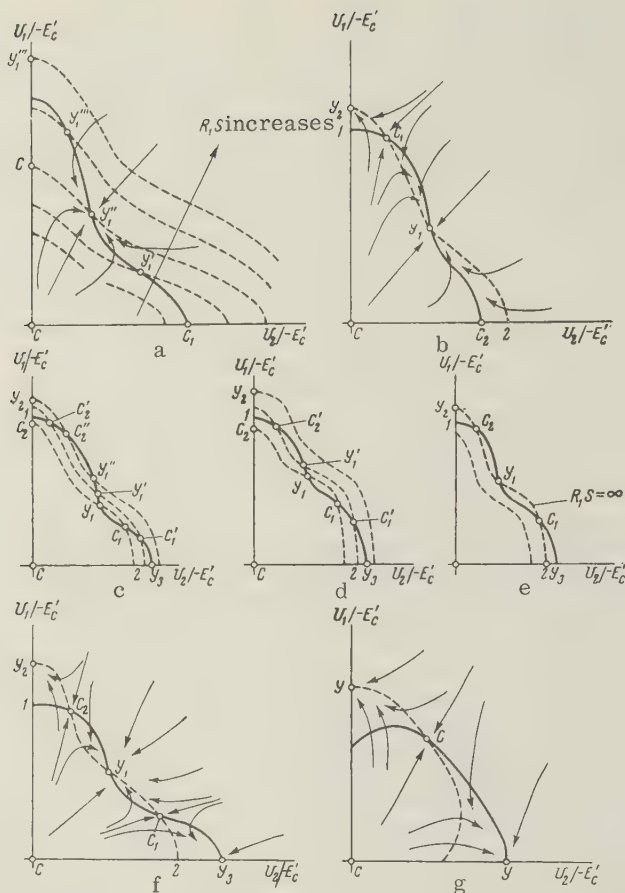


Figure 5. Location of singular points and behavior of integral curves of equation (1):

dashed line, solution of equation $S_1(x, y) = 1/R_1S$; solid line, solution of equation $S_2(x, y) = 1/R_2S$; Y, singular point of node type; C, singular point of saddle type.

2. $R_2S=5$. Qualitative distribution of the principal isoclines is shown in Figure 5b. The nature of the change in the dashed isocline is the same as in Figure 5a, that is, with an increase in R_1S it expands from the coordinate origin. However, in this case there appear two intersections at the center (points C_1 and Y in Figure 5b). Now it is not difficult to plot the amplitude of oscillation as a function of R_1S . At first it is the same as with $R_2S=3$ (Figure 6a). However, this will be the case until singular points C_1 and Y_1 in Figure 5b coincide (in Figure 6b this corresponds to $\log R_1S=1.0$). At this moment there occurs extinction of oscillation in the second tuned circuit and the amplitude of oscillation in the first tuned circuit increases to the amplitude of free oscillation (the jump at point Y_2 in Figure 5b). With a further increase in R_1S ($\log R_1S > 1$) free oscillations occur in the first tuned circuit. If R_1S is now decreased, free oscillation will occur in the first tuned circuit until points Y_2 and 1 in Figure 5b coincide; in this case ($\log R_1S=0.9$) there occurs a jump from point Y_2 to point Y_1 . As a result there is observed a sharp drop in amplitude of oscillation in the first tuned circuit and the commencement of oscillation in the second. Thus, we see that, in distinction from the case of undervoltage operation ($R_2S=3$), in overvoltage operation (critical operation with $\log R_1S=0.82$) hysteresis phenomena are observed.

3. $R_2S=8, 9, 25$. This is the most complex case. The oscillator is constantly in the overvoltage state. The corresponding amplitude plane is represented in Figure 5f. The

number of singular points is now six, three of which (Y_1 , Y_2 , and Y_3) are stable and three of which (C , C_1 , and C_2) are unstable.

Depending on the "microstructure" of the principal isoclines, three subcases of amplitude behavior with a change in R_1S are possible. We shall examine these individually.

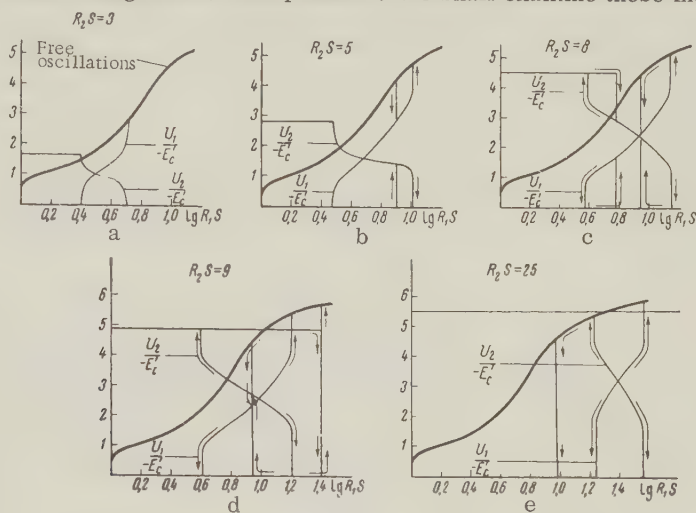


Figure 6. Amplitude of oscillation as a function of regeneration of the first tuned circuit.

3a. $R_2S = 8$. The dynamics of change in principal isoclines are shown in Figure 5c and the amplitude curves are shown in Figure 6c. As long as the solid curve embraces the dashed curve oscillations exist only in the second tuned circuit (point Y_3). With an increase in R_1S the dashed curve expands and there appear two (C_1 , Y_1) and then three (C' , Y'_1 , C'_2) points of intersection of the dashed curve with the solid isocline. However, until points 2 and Y_3 coincide oscillation occurs only in the second tuned circuit. The microstructure of the principal isoclines is such that at the moment of coincidence of points 2 and Y_3 there exist two intersections (points Y'' , C'') of the dashed and solid isoclines; hence, from point Y_3 there occurs a jump to point Y'_1 (with $\log R_1S = 0.78$ in Figure 6c). This jump indicates a sharp drop in amplitude of oscillation in the second tuned circuit and the appearance of oscillation in the first tuned circuit. A further increase in R_1S (from the value $\log R_1S = 0.78$) leads to coincidence of points Y'_1 and C'_2 (at $\log R_1S = 1.14$ in Figure 6c) and a jump to point Y_2 in Figure 5c. Moreover, free oscillations occur only in the first tuned circuit. If (at $\log R_1S > 1.4$) we begin to decrease R_1S , these oscillations continue until the moment of coincidence in Figure 5c of points Y_2 and 1 (in Figure 6c with $\log R_1S = 0.94$) and the jump to point Y_1 . As a result, oscillations occur in the second tuned circuit and there is a sharp decrease in amplitude of oscillation in the first tuned circuit. A further decrease in R_1S (from $\log R_1S = 0.94$) leads to coincidence in Figure 5c of points Y_1C_1 (at $\log R_1S = 0.57$) and a jump to point Y_3 (oscillations occur only in the second tuned circuit). Thus, in this subcase (Figure 6c) there are two regions in which hysteresis phenomena are observed.

3b. $R_2S = 9$. The specific dynamics of the isoclines for this case are shown in Figure 5d and the amplitude characteristics are given in Figure 6d. Oscillations exist in the second tuned circuit until, as in Figure 5d, points 2 and Y_3 coincide (at $\log R_1S = 1.4$ in Figure 6d). The microstructure of the principal isoclines is such that by this moment there are no intersections at the center (the upper dashed curve in Figure 5d); hence, there occurs a jump directly to point Y_2 . This indicates that oscillation in the second tuned circuit ceases and in the first tuned circuit attains the amplitude of free oscillation (see Figure 6d). If R_1S is further increased ($\log R_1S > 1.4$), then, as in subcase 3a ($R_2S = 8$), oscillation occurs only in the first tuned circuit and, with a decrease in R_1S , there at first arise double-frequency oscillations (in Figure 6d, with $\log R_1S = 0.94$) and then single-frequency oscillations in the second tuned circuit (in Figure 6d, with $\log R_1S = 0.6$). If upon decreasing regeneration R_1S to the value $\log R_1S = 0.94$ (the moment of occurrence of double-frequency oscillation) it then begins to

increase again, at $\log R_1 S = 1.2$ (Figure 6d) in Figure 5d there occur coincidence of points C'_2 and Y'_1 , a jump to point Y_2 , and single-frequency oscillation in the first tuned circuit.

3c. $R_2 S = 25$ (Figure 5e and 6e). In this subcase the structure of the principal isoclines is such that in Figure 5e points 2 and Y_3 do not coincide even at $R_1 S = \infty$. This indicates that

oscillation in the second tuned circuit cannot be extinguished by any change in $R_1 S$ (this condition corresponds to the horizontal straight line U_2/E'_0 in Figure 6e). If oscillations exist in the first tuned circuit (point Y_2), then with a decrease in $R_1 S$ coincidence of points Y_2 and 1 occurs and there is a jump directly to point Y_3 , that is, oscillations occur only in the second tuned circuit since the microstructure of the principal isoclines is such that at the moment of coincidence of Y_2 and 1 intersections no longer occur in the central portion of the isoclines (in Figure 5e, points C_1 and C_2).

Thus, we conclude that if single-frequency oscillations exist in any of the tuned circuits in a system, it is not possible in this case to obtain double-frequency oscillations by changing $R_1 S$. However, in this case there is a region in which double-frequency oscillation occurs ($1.24 < \log R_1 S < 1.56$ in Figure 6e). If we are dealing with this region, the behavior of the amplitudes with a change in $R_1 S$ will be the same as with double-frequency oscillation in the case in Figure 6c.

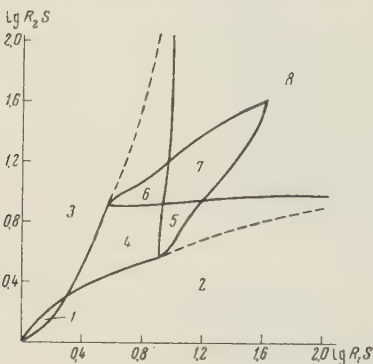


Figure 7. Segmentation of parameter plane according to the nature of oscillations.

4. Figure 5g. $R_2 S = 75$. The number of singular points is decreased to four. Upon coincidence (Figure 5f) of points C_1 , and Y_1 and C_2 points C_1 and C_2 impart to point Y_1 its instability and within the system there is possible only single-frequency oscillations in one or the other tuned circuit, depending on the initial conditions.

Now, by using the curves of Figure 6a-6e, we may perform segmentation of the parameter ($\log R_1 S$, $\log R_2 S$) according to the nature of the oscillations (Figure 7). For this purpose let us choose a definite value of $R_2 S$ and in the parameter plane (Figure 7) let us draw a horizontal line at the corresponding level. Using Figures 6a-6e, let us divide this line into segments corresponding to the different nature of oscillations. By joining the ends of the corresponding segments, we obtain the segmentation of the parameter plane.

The results of this segmentation are shown in Figure 7. Several regions are apparent therein: 1) single-frequency oscillation in the first or second tuned circuit, depending on the initial conditions; 2) oscillations in the first tuned circuit; 3) oscillations in the second tuned circuit; 4) double-frequency oscillations; 5) double-frequency oscillations or oscillations in the first tuned circuit, depending on the initial conditions; 6) double-frequency oscillations or oscillations in the second tuned circuit, depending on the initial conditions; 7) oscillations in either of the tuned circuits or double-frequency oscillations, depending on the initial conditions; 8) single-frequency oscillations in either of the tuned circuits, depending on the initial conditions.

For the sake of comparison, Figure 7 includes a dashed curve indicating segmentation of the parameter plane in the case where the overvoltage condition is not considered. We see that with consideration of overvoltage operation the region in which double-frequency oscillation is possible is finite in value and, in addition, is broken down into several subregions (4-7), of which only subregion 4 is marked by unconditional existence of double-frequency oscillations.

This report was prepared under the direction of S. I. Yevtyanov, to whom the author expresses his sincere thanks for great assistance in the research and in reviewing the manuscript.

REFERENCES

1. S. I. Yevtyanov, Oscillator calculations in overvoltage operation, *Elektrosvyaz'*, 1957, 11, 52.
2. A. N. Bruyevich, Asynchronous oscillation in a self-excited oscillator with two degrees of freedom, *Radiotekhnika i Elektronika*, 1960, 5, 10, 1559.

Submitted to the editors 7 July 1960

DECOUPLING OF TWO CLOSELY-SPACED REFLECTORS

B. Ye. Kinber

Different types of coupling between large reflectors (between radiators, through one rim, through two rims) are analyzed on the basis of the geometric theory of diffraction.

Formulas are derived for calculation of different types of coupling and the frequency relationships of these couplings are analyzed.

INTRODUCTION

The transition loss η (or coupling coefficient) between closely-spaced antennas is of great practical importance in all fields of application of antennas in the microwave range. Unfortunately, this factor has been investigated only in dipole antennas (reference [1]), that is, antennas with low directivity.

The present report discusses the coupling between closely spaced reflectors, the dimensions of which are considerably greater than a wavelength and also the coupling between radiators separated by reflectors of finite dimensions. It is assumed that the antennas are so located that their "principal" rays are not aligned and the coupling between antennas is a weak and stray phenomenon.

The present report does not consider the strong coupling existing when a large part of the energy radiated by the transmitting antenna (wireless radio transmission lines, periscopic antennas) must be received by the receiving antenna.

1. GENERAL FORMULA FOR COUPLING BETWEEN ANTENNAS

As was shown in reference [2], if we do not consider multiple reflections (diffraction) of the field between the antennas, the power received by antenna I (P_{rcI}) may be expressed in terms of the field radiated by antenna II and incident at antenna I ($E_{II}H_{II}$) and the field radiated by antenna I during transmission operation ($E_I H_I$):

$$P_{rcI} = \frac{\left| \int_{s_1} \langle \vec{n}, \{ [\vec{E}_I \vec{H}_{II}] - [\vec{E}_{II} \vec{H}_I] \} \rangle ds \right|^2}{4 \int_{s_2} \langle \vec{n}, \{ [\vec{E}_I^* \vec{H}_I] + [\vec{E}_{II} \vec{H}_I^*] \} \rangle ds}, \quad (1)$$

where s_1 is a closed surface around antenna I (aperture); s_2 is the surface over which the total energy flux is calculated during transmission operation of antenna I. Surface s_2 may be chosen arbitrarily (e.g., transverse to the line of transmission or so that it coincides with s_1).

The quantity in which we are interested, η , the coefficient of coupling between antennas, is the ratio of (1) and the total power radiated by antenna II:

$$\eta = \frac{P_{rcI}}{P_{trII}} = \frac{\left| \int_{s_1} \langle \vec{n}, \{ [\vec{E}_I \vec{H}_{II}] - [\vec{E}_{II} \vec{H}_I] \} \rangle ds \right|^2}{16 P_{trI} P_{trII}}, \quad (2)$$

where

$$P_{\text{tr I}} = \frac{1}{4} \int_{s_2} (\vec{n}, \{[\vec{E}_I^* \vec{H}_I] + [\vec{E}_I \vec{H}_I^*]\}) ds; \quad (3)$$

$$P_{\text{tr II}} = \frac{1}{4} \int_{s_3} (\vec{n}, \{[\vec{E}_{II}^* \vec{H}_{II}] + [\vec{E}_{II} \vec{H}_{II}^*]\}) ds$$

is the total power radiated by antennas I and II; s_3 is the surface equivalent to s_2 but for transmitting antenna II.

Although formula (2) is derived only for single diffraction at antenna I, it is shown below that it may also be used in considering multiple diffractions between antennas.

Let us now describe the near field of the antennas in the form of the sum of terms satisfying the Fermat principle (references [3, 4]), that is, terms whose phases are proportional to the extreme paths from the antenna radiator to the point of integration at s_1 :

$$\begin{aligned} \vec{E}_I &= \sum_n \vec{f}_n e^{ik\psi_n}, \\ \vec{H}_I &= \sum_n \vec{\Phi}_n e^{jk\psi_n}, \\ \vec{E}_{II} &= \sum_m \vec{f}_m e^{ik\psi_m}, \\ \vec{H}_{II} &= \sum_m \vec{\Phi}_m e^{ik\psi_m}, \end{aligned} \quad (4)$$

and insert (4) into (2).

The common term of the resulting double sum has the form

$$\int_{s_1} (\vec{n} \{[\vec{f}_n \vec{\Phi}_m] - [\vec{f}_m \vec{\Phi}_n]\}) e^{ik(\psi_n + \psi_m)} ds. \quad (5)$$

Its amplitude is a slowly varying function and the phase has a maximum point. Consequently, for calculation of the integral we may use the stationary phase method.

It is easily seen (Figure 1) that the maximum $\psi_n + \psi_m$ at surface s_1 corresponds to the maximum path between the radiators of antennas I and II and that the link of this maximum path joining antennas I and II is a segment of line Q intersecting surface s_1 at point Y*. In the calculations we shall choose surface s_1 individually for each term of the series, wherein s_1 will be perpendicular to Q at a point (along the line) of stationary phase. Such choice of s_1 is convenient in that at a point (of the line) of stationary phase $\nabla\psi_n$ and $\nabla\psi_m$ are collinear with Q and the vectors \vec{f} and $\vec{\Phi}$ are perpendicular to one another and lie on surface s_1 .

$$\begin{aligned} \vec{\Phi}_m &= -\frac{1}{\Omega} [n \vec{f}_m], \\ \vec{\Phi}_n &= \frac{1}{\Omega} [n \vec{f}_n], \end{aligned} \quad (6)$$

where $\Omega = 120\pi$ is the total resistance of free space and, consequently,

$$(\vec{n} \{[\vec{f}_n \vec{\Phi}_m] - [\vec{f}_m \vec{\Phi}_n]\}) = -\frac{2}{\Omega} (\vec{f}_n \vec{f}_m). \quad (7)$$

In subsequent calculations we may limit ourselves to the first term of expansion according to the stationary phase method. With this in mind (2) may be written in the form

$$\begin{aligned} \eta &= \frac{1}{4P_{\text{tr I}} P_{\text{tr II}} \Omega^2} \left| \sum_n \sum_m (\vec{f}_n \vec{f}_m)_{\text{st}} e^{ik(\psi_n + \psi_m)_{\text{st}}} \times \right. \\ &\quad \left. \times \int_{s_1} \exp \{ik[\psi_n - \psi_{n \text{ st}} + \psi_m - \psi_{m \text{ st}}]\} ds \right|^2. \end{aligned} \quad (8)$$

*As is shown below, there may be cases in which $\psi_n + \psi_m$ yield not a point but a line of stationary phase on s_1 .

2. TYPES OF COUPLING BETWEEN ANTENNAS

As shown above, the formula for coupling between antennas contains the sum of terms, the phase of which is proportional to the maximal paths between antennas. Hence, we shall subsequently say more simply that the coupling between antennas is achieved over various maximum paths. In this section we shall examine the more important types of maximum paths and show how multiple diffractions may be taken into account.

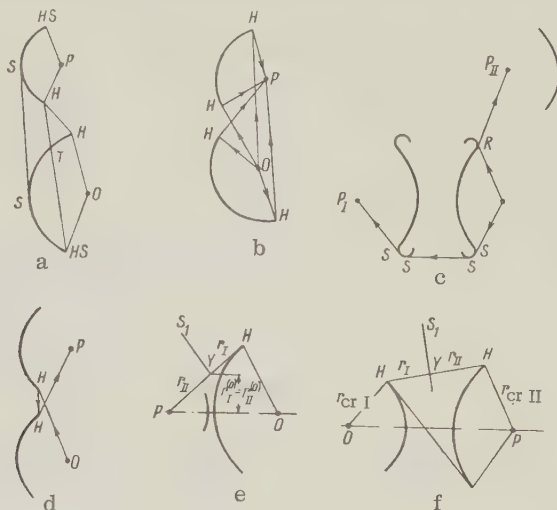


Figure 2. Possible types of coupling between antennas.

The maximum path between the radiators of antennas I and II is a broken line consisting of segments of straight lines and arcs belong to the antennas. Part of the links belong to one antenna, part belong to the other, and, finally, there is a common link joining the two antennas.

Each maximum path is characterized by a number of points of stationary phase of type R, H, S, and T (see reference [15]) at one and the other antenna. For example, the path O-H-H-P (Figure 2a) consists of segment O-H from radiator O to point H on the rim of one antenna, the similar link H-P at the other antenna and the common link H-H. The path O-HS-S-S-SH-P consists of link O-HS from the radiator to the rim of the first antenna, the link HS-S along the rear surface from the rim HS to the point of reflection S at the first antenna, similar links at the second antenna, and the common link S-S.

The indicated chains (e.g., O-H-H-P, O-H-P) do not uniquely characterize the maximum paths. Thus, there are four extreme paths O-H-P (see Figure 2b) and 12 maximum paths O-H-H-P. Part of those paths may be shadowed (e.g., one of the paths O-H-H-P in Figure 2a), that is, along with the maximum path O-H-H-P there exists a maximum path O-H-T-H-P.

We shall not deal with such paths here, for, in virtue of the condition of shadowing, they may be disregarded [4].

The types of coupling between antennas depend both on the type of antenna and on their relative orientation.* For example, the number of maximum paths O-H-H-P may vary from 1 to 12, depending on the relative orientation of the reflectors. For reflectors with rounded rim the path O-H-P is replaced by O-R-P or O-S-S-P (Figure 2c).

The above-mentioned extreme paths have corresponded to the maximum $\psi_n + \psi_m$, have contained only one common link between antennas and have defined any term of series (8).

However, maximum paths are possible which contain not one but several common links between antennas and which correspond not to single but to triple, quintuple, etc. diffractions. The simplest maximum path of this type is the path O-H-H-P, shown in Figure 2d (three

*However, the amount of coupling in any case is infinite (to any maximum path there may be added the closed maximum path around each of the antennas, consisting of any whole number of revolutions).

common links) and the corresponding triple diffraction between antennas. It is evident that its contribution to coupling may not prove to be less than that from the above-mentioned paths O-H-H-P corresponding to single diffractions.

Consideration of such multiple diffractions is necessary, since it is generally difficult to determine whether the object of analysis is to be treated as two interconnected antennas or as one complex antenna, while the values of coupling in single and multiple diffraction may prove commensurate.

Since there is no difference in description of diffraction of the primary wave of a radiator at its "own" or a "foreign" rim, multiple diffractions may be introduced into the calculation by entering the corresponding terms into sum (8).

Finally, let us discuss the maximum paths for axially symmetrical and coaxial antennas. In this case the maximum paths — the path O-H-P (Figure 2e), two paths O-H-H-P (Figure 2f) and more complex paths — are conical surfaces. Consequently, coupling in these cases occurs not through one point of the rim, but through the entire rim of the reflector. It is natural that here there will be observed an increase in coupling.

The total coupling coefficient η is defined by the square of the sum of maximum terms (8) and hence the coefficients of coupling from the individual terms are not additive and the dependence of η on wavelength, antenna dimensions and orientation must be nonmonotonic. In practice the value of coupling is determined from one or several of the largest terms of sum (8).

3. CHARACTERISTICS OF THE PRINCIPAL TYPES OF COUPLING

Let us conduct a qualitative examination of the more important forms of coupling: O-P, O-H-P, O-H-H-P. The last two couplings we shall examine both for the case of a stationary point and for the case of a stationary line. In the analysis we shall assume that the investigated coupling is the determining one and all others may be disregarded.

Let us perform analysis for axially symmetrical reflectors of radius R with sharp edge.

It is assumed that the axes of the reflectors lie in one plane. The maximum paths lie in this plane and the planes of the principal curvatures of the wave surfaces of fields I and II coincide and lie both in this plane and the plane perpendicular to it. The position of surface s_1 does not affect the results of calculation and may be expressed arbitrarily.

Coupling O-P (Coupling of Radiators). The field of the primary radiators of the antennas at point of stationary phase Y on s_1 (Figure 3)

$$\begin{aligned}\vec{E}_I &= \sqrt{60P_{\text{tr}} G_I} \frac{\vec{T}_I(\varphi_I, \psi_I)}{r_I} e^{ikr_I}, \\ \vec{E}_{II} &= \sqrt{60P_{\text{tr}} G_{II}} \frac{\vec{T}_{II}(\varphi_{II}, \psi_{II})}{r_{II}} e^{ikr_{II}},\end{aligned}\quad (9)$$

where φ_I, φ_{II} are the angles between the planes of the origin of angle measurement and the plane of coupling formed by the antenna axes; ψ_I, ψ_{II} are the angles between the antenna axes and the line OP; G_I and G_{II} are the directivity coefficients of the antenna radiators; the radii of curvature of the waves

$$\rho_I^{(1)} = \rho_I^{(2)} = r_I, \quad \rho_{II}^{(1)} = \rho_{II}^{(2)} = r_{II}. \quad (10)$$

Hence for the investigated coupling O-P expression (8) has the form

$$\begin{aligned}\eta &= \frac{1}{4P_{\text{tr}} I P_{\text{tr}} II (420\pi)^2} \left| 60 \frac{\sqrt{P_{\text{tr}} I P_{\text{tr}} II G_I G_{II}}}{r_I r_{II}} (\vec{T}_I \vec{T}_{II}) \int_{s_I} \exp \left\{ \frac{ik}{2} \left(\frac{1}{r_I} + \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1}{r_{II}} \right) (x^2 + y^2) \right\} dx dy \right|^2 = \frac{\lambda^2 G_I G_{II} (\vec{T}_I \vec{T}_{II})^2}{(4\pi L_{OP})^2},\end{aligned}\quad (11)$$

Figure 3. For calculation of coupling O-P.

cides with point Y.

The derived formula shows that coupling between the radiators satisfies the ordinary coupling equation. This obvious conclusion confirms the validity of the above calculations.

Coupling O-H-P (Stationary Point). For calculation of this coupling, in addition to the field of the primary radiator, it is necessary to have an expression for the field of the fringe wave.

For normal incidence of the wave at the edge of the semiplane the expression for the fringe wave may be presented in the form (reference [6])

$$\vec{E}_{cr} = \sqrt{\frac{2}{\pi k r}} e^{i(kr + \frac{\pi}{4})} \{ \vec{\varphi} E_{0\varphi} F_{\varphi}(\theta_0, \theta) + \vec{\theta} E_{0\theta} F_{\theta}(\theta_0, \theta) \}, \quad (12)$$

where (see Figure 4) r is the distance from the rim to the point of observation; $\vec{\varphi}$ is the unit vector parallel to the rim; $\vec{\theta}$ is the unit vector perpendicular to the rim and vector \vec{r} ; $E_{0\varphi}$, $E_{0\theta}$ are the φ - and θ -components of the primary field at the rim;

$$F_{\varphi} = \frac{\cos \frac{\theta_0}{2} \sin \frac{\theta}{2}}{\cos \theta - \cos \theta_0}; F_{\theta} = \frac{\sin \frac{\theta_0}{2} \cos \frac{\theta}{2}}{\cos \theta - \cos \theta_0} \quad (13)$$

give the directivity patterns of the fringe wave for the φ - and θ -components; θ_0 and θ are the angle of incidence and the angle of observation.

The above definition of the φ - and θ -components differs from the standard definition but is suitable for the calculations given below.

It is easily shown* that the expression for the fringe wave of an axially symmetrical reflector differs from (12) by the replacement of the amplitude factor for a cylindrical wave $1/\sqrt{r}$ (linear current filament) with the factor for a toroidal wave $1/\sqrt{R/r^{(0)}r}$ (annular current filament):

$$\vec{E}_{cr} = \sqrt{\frac{2R}{\pi k r^{(0)} r}} e^{i(kr + \frac{\pi}{4})} \{ \vec{\varphi} E_{0\varphi} F_{\varphi} + \vec{\theta} E_{0\theta} F_{\theta} \}, \quad (14)$$

where r is the distance from stationary point H at the rim to the observation point; $r^{(0)}$ is the distance from the observation point to the reflector axis (it is taken with a plus sign if the observation point and stationary point H lie on the same side of the axis and with a minus sign if these points lie on different sides of the axis).

By the φ -component is meant the component perpendicular to the plane of reflection (meridional plane) and by the θ -component is meant the component lying in this plane and perpendicular to r .

The radii of curvature of the wave are

$$\rho^{(1)} = r, \quad \rho^{(2)} = \frac{|r^{(0)}|}{\sin \alpha}, \quad (15)$$

where α is the angle between \vec{r} and the axis of the annulus.

Now let us determine the transition loss for coupling O-H-P shown in Figure 5. On the principle of reciprocity it does not matter whether point H is at the receiving or transmitting antenna.

Inserting expressions (9), (14) and (15) into (8), we obtain

$$\begin{aligned} \eta &= \left| \frac{15 \sqrt{G_I G_{II}}}{120\pi} \{ T_{\varphi I} F_{\varphi I} T_{\varphi II} + T_{\theta I} F_{\theta I} T_{\theta II} \} \sqrt{\frac{2R_I}{\pi k r_I^{(0)} r_I}} \times \right. \\ &\times \int_{-\infty}^{+\infty} \exp \left\{ -\frac{ik}{2} \left[\left(\frac{1}{r_I} + \frac{1}{r_{II}} \right) x^2 + \left(\frac{1}{r_{II}} + \frac{\sin \alpha_I}{|r_I^{(0)}|} \right) y^2 \right] \right\} dx dy \Big|^2 = \\ &= \frac{\lambda^3 G_I G_{II} (T_{\varphi I} F_{\varphi I} T_{\varphi II} + T_{\theta I} F_{\theta I} T_{\theta II})^2 \sin^2 \psi_{cr}}{16\pi^4 R_I L_{OH} S_I}, \end{aligned} \quad (16)$$

*The lines consisting of the caustic of the toroidal wave of the current annulus and the axis of the annulus. In their vicinity the field must have the nature of a cylindrical wave, that is, it must be proportional to $1/\sqrt{r}$ and $1/\sqrt{r^{(0)}}$. The subsequent formula satisfies the presented conditions. Its validity may also be checked by direct calculation.

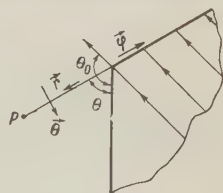


Figure 4. For determination of the geometric parameters of wave diffraction at the semiplane.

where $L_{OH} = r_I + r_{II}$ is the distance from the rim to the "foreign" radiator; $S_I = |r_I(0)| + \sin \alpha_I r_{II}$ is the distance from radiator O to the axis of the other antenna.

The parameters (distance and angles) entering into formula (16) are determined from the geometry of the antennas and their relative positions. The influence of a shield interposed between the two antennas and two radiators is also described in the above formula.

Coupling O-H-P (Stationary Line). If two axially symmetrical antennas are placed back to back and are coaxial, coupling occurs over the entire rim (Figure 2f). This is also true in the case of a circular shield the axis of which coincides with the line joining the radiators.

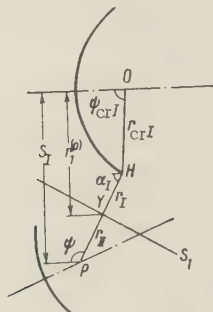


Figure 5. For calculation of coupling O-H-P.

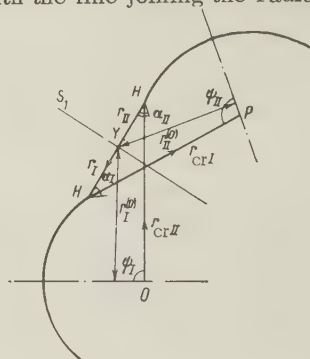


Figure 6. For calculation of coupling O-H-H-P.

As surface s_1 it is convenient to choose a cone perpendicular to the beam of rays from the radiator to the rim (Figure 2e).

In calculating this coupling it is necessary to know the polarization characteristic of the radiator. Let us give it in the form (reference [5])

$$\vec{T}(\varphi, \theta) = -(\vec{\theta} \cos \varphi + \vec{\varphi} \sin \varphi) T(\theta). \quad (17)$$

Inserting expressions (9), (14), (15) and (17) into (8) and considering that the stationary phase method must be used only in integrating in the meridional plane and that the line of stationary phase is removed from the axis by a distance $r_I^{(0)}$, we obtain

$$\eta = G_I G_{II} \left| \frac{1}{8\pi} \sqrt{\frac{2R_I}{\pi k r_I^{(0)} r_I}} \frac{r_I^{(0)}}{r_{crI} r_{II}} T_I(\psi_{crI}) T_{II}(\psi_{II}) \int_{-\infty}^{+\infty} \exp \left[\frac{ik}{2} \left(\frac{1}{r} + \frac{1}{r_{II}} \right) t^2 \right] dt \times \right. \\ \left. \times \int_0^{2\pi} [F_{\theta I} \cos^2 \varphi + F_{\varphi I} \sin^2 \varphi] d\varphi \right|^2 = \frac{\lambda^2 G_I G_{II}}{128\pi^4} (F_{\theta I} + F_{\varphi I})^2 \frac{\sin^2 \psi_{crI}}{L_{OH} r_I}, \quad (18)$$

since $R_I/r_{crI} = \sin \xi_{crI}$.

In calculating the influence of a shield placed between noncoaxial radiators but with a cophasal shield rim, the expression for the rim integral has a more complex form, but the dependence on λ and shield dimensions remains the same.

Coupling O-H-H-P (Stationary Points). Coupling O-H-H-P exists in twelve forms. Calculation and dependence on wavelength is identical for all. Hence, as an example, let us discuss coupling through two closely spaced rims for triple diffraction. The geometry of this coupling is shown in Figure 6.

Inserting the expression for fringe waves (14) into (8), we obtain

$$\eta = G_I G_{II} \left| \frac{1}{8\pi} \sqrt{\frac{2R_I}{\pi k r_I^{(0)} r_I}} \sqrt{\frac{2R_{II}}{\pi k r_{II}^{(0)} r_{II}}} \frac{1}{r_{crI} r_{crII}} \times \right. \\ \left. \times \{T_{\varphi I} F_{\varphi I} T_{\varphi II} F_{\varphi II} + T_{\theta I} F_{\theta I} T_{\theta II} F_{\theta II}\} \int_{-\infty}^{+\infty} \exp \left\{ \frac{ik}{2} \left[\left(\frac{1}{r_I} + \frac{1}{r_{II}} \right) x^2 - \right. \right. \right. \\ \left. \left. \left. + \left(\frac{\sin \alpha_I}{|r_I^{(0)}|} + \frac{\sin \alpha_{II}}{|r_{II}^{(0)}|} \right) y^2 \right] \right\} dx dy \right|^2 = \frac{\lambda^4 G_I G_{II} \sin^2 \psi_{crI} \sin^2 \psi_{crII}}{16\pi^6 R_I R_{II} L_{HH} S_{II}} \times \{T_{\varphi I} F_{\varphi I} T_{\varphi II} F_{\varphi II} + T_{\theta I} F_{\theta I} T_{\theta II} F_{\theta II}\}^2, \quad (19)$$

where

$$S_{II} = \sin \alpha_I R_{II} + \sin \alpha_{II} R_I + \sin \alpha_I \sin \alpha_{II} L_{HH};$$
$$L_{HH} = r_I + r_{II}.$$

Coupling O-H-H-P (Stationary Line). The geometry of this coupling for the case of back-to-back coaxial placement of the antennas (one of the twelve variants) is shown in Figure 2f. In calculation, as for coupling O-H-P, we take the radiation polarization structure given in (17). The aperture-cone surface is perpendicular to the ray cone H-H.

$$\eta = G_I G_{II} \left| \frac{1}{8\pi} \sqrt{\frac{2R_I}{\pi k r_I^{(0)} r_I}} \sqrt{\frac{2R_{II}}{\pi k r_{II}^{(0)} r_{II}}} \frac{r_I^{(0)} T_I T_{II}}{r_{cr I} r_{cr II}} \int_{-\infty}^{+\infty} \exp \left\{ \frac{ik}{2} \left(\frac{1}{r_I} + \frac{1}{r_{II}} \right) t^2 \right\} dt \times \right. \\ \left. \times \int_0^{2\pi} \{ F_{\varphi I} F_{\varphi II} \sin^2 \varphi + F_{\theta I} F_{\theta II} \cos^2 \varphi \} d\varphi \right|^2 = \frac{\lambda^3 G_I G_{II} T_I^2 T_{II}^2 \sin^2 \psi_{cr I} \sin^2 \psi_{cr II}}{428 \pi^6 R_I R_{II} L_{HH}}. \quad (20)$$

4. COMPARISON OF DIFFERENT COUPLINGS

The above analysis permits us to compare the individual types of coupling. The principal relationships are listed in Table 1.

Table 1

№ п. п.	Type of coupling	Depen- dence on λ	Dependence on distance be- tween antennas L	Dependence on diameter of reflector R
1	O—P	λ^3	$(L_{OP})^{-2}$	—
2	O—H—P, stationary line	λ^3	$(L_{OH})^{-1}$	R^{-1}
3	O—H—P, stationary point	λ^3	$(L_{OH} S_I)^{-1}$	H^{-1}
4	O—H—H—P, stationary line	λ^3	$(L_{HH})^{-1}$	$(R_I R_{II})^{-1}$
5	O—H—H—P, stationary point	λ^4	$(L_{HH} S_{II})^{-1}$	$(R_I R_{II})^{-1}$

It follows from Table 1 that the strongest coupling is that between the radiators (O-P). Unless special measures are adopted (for example, decreasing the side lobes of the radiator, polarization decoupling of the radiators or shielding them from one another), this coupling is the principal coupling and determines the transition loss. Since coupling O-P is only one and the other types of coupling are of smaller order, the frequency dependence of transition loss is monotonic. If the antennas are turned away from one another and the radiators shielded from one another, the determining coupling is the coupling across one rim O-H-P and is smaller by one order of magnitude. If the rim is cophasal relative to the radiator (the contour of the reflector or shield is located at the surface of an ellipsoid of rotation whose foci are located at points O and P), the order of coupling is the same as that of O-P. The multiplicity of couplings O-H-P with different lengths of path leads to a nonmonotonic dependence of transition loss on frequency. The value of coupling O-H-P depends considerably on the radiation polarization.

Weaker coupling O-H-H-P, appearing only in the absence of stronger couplings, is characteristic for antennas of approximately the same diameter and placed back to back. As in the case of coupling O-H-P, coaxial placement of the antennas considerably increases the coupling between them.

Let us evaluate the order of magnitude of coupling in the example of two antennas with diameter of 20λ and angle of radiation 120° . The level of illumination of the reflector edge, as usual, is taken to be 0.3. The directivity coefficient of the radiator in this case is equal to 4. The side-lobe level of the radiators is assumed equal to 0.2. Polarization of the antenna is assumed to be identical and the plane of polarization to coincide with the plane of

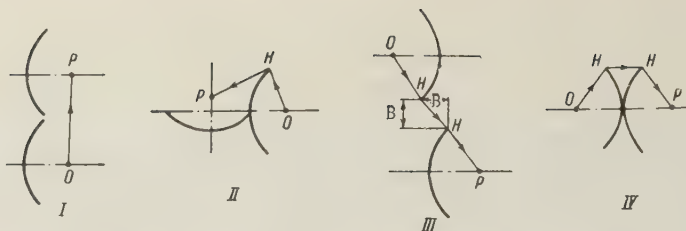


Figure 7. Various antenna positions (for sample calculation), $B = 5 \lambda$.

coupling. Let us examine the following variants of antenna placement.

1. The antennas are oriented parallel to one another and their edges touch (Figure 7, I).
2. The antennas are oriented perpendicular to one another. One of the antennas is shielded from the other (Figure 7, II).
3. The antennas face in opposite directions and are axially displaced relative to one another (Figure 7, III).
4. The antennas are coaxial and placed back to back. The reflectors are in contact (Figure 7, IV).

Table 2 lists the values of coupling between the antennas as calculated for these variants. The geometric parameters required for calculation are determined in accordance with Figure 7.

In order of magnitude the calculated values of transition loss coincide with the measured values (references [7, 8]). Comparison of variants II and IV shows that coaxial placement of the antenna considerably decreases decoupling between them.

Table 2

Antenna position	Types of coupling	Transition loss
I	$O-P$	40
II	$O-H-P$, stationary point	70
III	$O-H-H-P$, stationary point	110
IV	$O-H-H-P$, stationary line	88

In conclusion, it is my pleasant duty to express my thanks to A. I. Shpuntov for his valuable discussions on the subject of this work.

REFERENCES

1. Antennas, translated from English under editorial supervision of A. I. Shpuntov, Izd. Sovetskoye Radio, 1951.
2. B. Ye. Kinber, Toward a receiving antenna theory, Radiotekhnika i Elektronika, 1961, 6, 4, 651.
3. Diffraction by an aperture, J. Appl. Physics, 1957, 28, 4, 426.
4. B. Ye. Kinber, Condition of shadowing and diffraction correction for current distribution, Radiotekhnika i Elektronika, 1960, 5, 9, 1407.
5. B. Ye. Kinber, Lateral Radiation of reflectors, Radiotekhnika i Elektronika, 1961, 6, 4, 545.
6. L. A. Vaynshteyn, Difraktsiya elektromagnitnykh i zvukovykh voln na otkrytom kontse volnovoda [Diffraction of electromagnetic and acoustic waves at the open end of a waveguide], Izd. Sovetskoye Radio, 1953.
7. E. S. Harris, Electronics, 1957, 30, 6, 204 (see RZh Elektrotehnika, 1958 abstract No. 30315).

Submitted to the Editors 17 August 1960

THE THEORY OF RADIO SHOCK WAVES IN NONLINEAR TRANSMISSION LINES

R. V. Khokhlov

This article deals with the process of propagation of electromagnetic shock waves in nonlinear transmission lines. By contrast with previously published studies, the effect of attenuation on the shaping and spreading of the shock wavefront is analyzed and the possibility of using the transmission line as a harmonic oscillator and a sawtooth-voltage generator is considered. For the analysis, a method is used based on the low value of the nonlinearity and the losses in the transmission lines.

INTRODUCTION

The problem of the propagation of waves in nonlinear media has long attracted the attention of mathematicians, mechanical engineers and physicists. This is because the process of propagation of perturbations in gas and liquid is important in many fields of science and engineering, and gases and liquids are nonlinear media, (see, for instance, [1]). Recently radio-physicists and radio-engineers have been faced with this problem in connection with the increasing use of electromagnetic materials with nonlinear characteristics in radio engineering.

For convenience the problem of the propagation of waves in nonlinear media can be divided into two parts: 1) the propagation of waves under the conditions of a strongly dispersed medium; and 2) the propagation of waves under the conditions of weak dispersion. In the first case, the existing harmonic components of the wave do not actually interact with one another at all, or the interaction takes place between a small number of components with phase velocities of nearly the same size. In the second case, all the harmonic components of the signal interact with one another resonantly, establishing the conditions for deformation of the wave with a large number of spatial harmonics.

It is natural to analyze first the limiting cases — the case of strong dispersion, where an interaction between two harmonic components take place, and the case of no dispersion, where all the harmonic components of the wave interact strongly with one another. It is to this latter case that the present article is devoted.

The process of propagation of electromagnetic waves in nonlinear systems in the absence of dispersion has already been considered in articles [2, 3, 4]. There an analysis of the formation of shock waves is given and the structure of the wavefront is studied on the basis of rigorous methods which were developed in gas dynamics and which are suitable for systems with a large nonlinearity, but not for attenuating systems. The structure of the wavefronts is studied in [3] with the attenuation of the medium taken into account. There it is shown that shock waves may occur under the conditions of attenuation.

The present article is devoted to the process of the propagation of waves in systems which are slightly nonlinear, absorb energy weakly, and are entirely deprived of the property of

dispersion. The treatment is based on an approximate method which is a natural generalization of the method of the shortened-equation method to the case of systems of partial differential equations. A similar approximate method was already used in [5] to analyze the process of the propagation of waves in lines with parameters changing according to the equations for a traveling wave. Here the process of propagation of waves is studied in a line whose nonlinear parameter is the distributed capacitance.

I. FORMAL SOLUTION OF THE PROBLEM

In beginning the study of the process of propagation of electromagnetic waves in a transmission line with a nonlinear distributed capacitance, we shall assume that the energy losses in the line are caused by two distributed resistances, of which the first is connected in parallel to the capacitance, while the second is connected in series (1). For the time being, we shall assume for the sake of simplicity that the series resistance R is equal to zero and that the dissipation of energy takes place only through the leak G . In this case, it is possible to describe the behavior of the currents I , the voltages, U , and the charges Q in the line by the following system of equations:

$$-\frac{\partial I}{\partial z} = \frac{\partial Q}{\partial t} + GU, \quad -\frac{\partial U}{\partial z} = L \frac{\partial I}{\partial t} \quad (1)$$

Here L and G are the inductance and conductance of the line. Eliminating the current from these equations, we obtain the equation

$$\frac{\partial^2 U}{\partial z^2} - L \frac{\partial^2 Q}{\partial t^2} - LG \frac{\partial U}{\partial t} = 0. \quad (2)$$

At the left end or input of the line, the voltage is given:

$$\text{at } z = 0 \quad U = \Phi(t). \quad (3)$$

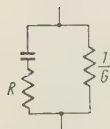


Figure 1

We shall take the line to the semiinfinitely in the z -direction and confine ourselves to looking for solutions of (2) having the form of waves traveling to the right. The question of the existence of these solutions is examined further on.

The charge Q on the capacitance is a nonlinear function of voltage. Since the further discussion is conducted for the case of a system which is only slightly nonlinear, the function $Q(U)$ can be taken to be

$$Q = CU + DU^2, \quad (4)$$

in the majority of cases, the coefficients C and D being coefficients in the expansion of the function $Q(U)$ into a series about the quiescent point. In addition, the variable part of the capacitance is assumed to be considerably smaller than its constant part C , i.e.,

$$DU_0 \ll C, \quad (5)$$

where U_0 is the amplitude of the voltage at the input. We also take the attenuation, i.e., the conductance G , to be small. Introducing the explicitly small parameter ϵ , we have $G = \epsilon G$, $D = \epsilon D$.

In the case where there is no attenuation or nonlinearity, equation (2) has a solution of the form

$$U = F(t - z\sqrt{LC}), \quad (6)$$

where F is an arbitrary function of the argument. It is natural to suppose that, in the case of small attenuation and nonlinearity, solution (2) is basically of the same form as (6), except that the function F changes slowly with increasing distance from the input, i.e.,

$$U = U(\epsilon z, t - z\sqrt{LC}). \quad (7)$$

Defining new variables (5)

$$\tau = t - z \sqrt{LC}, \quad \varepsilon z = x \quad (8)$$

and substituting them into Equation (2), we have

$$\varepsilon^2 \frac{\partial^2 U}{\partial x^2} - 2\varepsilon \sqrt{LC} \frac{\partial U}{\partial x} - \varepsilon LD \frac{\partial^2 U^2}{\partial \tau^2} - \varepsilon LG \frac{\partial U}{\partial \tau} = 0. \quad (9)$$

If we neglect the small terms, which are of the order of ε^2 , we obtain

$$\frac{\partial}{\partial \tau} \left[2 \sqrt{LC} \frac{\partial U}{\partial x} + LD \frac{\partial U^2}{\partial \tau} + LGU \right] = 0. \quad (10)$$

Integrating over τ , we obtain an equation which is considerably simpler than (2). This equation describes the process of the propagation of waves along the given nonlinear transmission lines:

$$\frac{\partial U}{\partial x} + \alpha U \frac{\partial U}{\partial \tau} + \delta U = F(x), \quad (11)$$

where $F(x)$ is generally an arbitrary function of x , which, in the case of boundary condition (3), should be set equal to zero. The parameters α and δ are expressed as follows:

$$\alpha = D \sqrt{\frac{L}{C}}, \quad \delta = \frac{1}{2} G \sqrt{\frac{L}{C}}. \quad (12)$$

A solution to Equation (11) with boundary condition (3) can be easily found. However, this solution generally expresses τ as a function of U and x :

$$\tau = \frac{\alpha}{\delta} (e^{\delta x} - 1) U + [\Phi]^{-1}(U e^{\delta x}), \quad (13)$$

where the symbol $[\Phi]^{-1}$ denotes the inverse function of Φ . In the case where there is no attenuation, this solution reduces to the form

$$\tau = \alpha x U + [\Phi]^{-1}(U) \quad \text{or} \quad t = z \left(\sqrt{LC} + U D \sqrt{\frac{L}{C}} \right) + [\Phi]^{-1}(U). \quad (14)$$

It is interesting to compare this solution with the exact solution of (2), which can be obtained for this case. It has the form

$$t = z \sqrt{L(C + 2D\bar{U})} + [\Phi]^{-1}(U). \quad (15)$$

By comparing (14) and (15) and remembering condition (5), we can easily establish the nature of the approximation of the method we are using.

$$\frac{dI}{dU} = \sqrt{\frac{C}{L}} \left(1 + \frac{D}{C} U + \frac{G}{2C} \frac{U}{\partial I} \right), \quad (16)$$

The change of the shape of the wave with distance can be analyzed by means of a simple graphical construction, which we carry out below.

2. RADIO SHOCK WAVES

Let us specify that the input voltage is sinusoidal:

$$z = 0 \quad U = U_0 \sin \omega t. \quad (17)$$

Substituting the corresponding inverse function into equation (13), we have

$$\omega \tau = U_0 \frac{\alpha \omega}{\delta} (1 - e^{-\delta x}) \frac{U e^{\delta x}}{U_0} + \arcsin \frac{U e^{\delta x}}{U_0}. \quad (18)$$

In beginning a graphical analysis of this equation, let us mark the value $\frac{U}{U_0} e^{\delta x}$, on the abscissa, and a value ω on the ordinate (Figure 2). According to (18), the solution can be

represented as the sum of two functions — the arcsine and a straight line with an angular coefficient Z :

$$Z = \frac{\omega \alpha U_0}{\delta} (1 - e^{-\delta x}). \quad (19)$$

After rotating the axis in Figure 2 and analyzing the form of the solution for different values of Z , we can see the distortion of the wave's profile at increasing distance from the input of the line (Figure 3).

The angular coefficient Z has the meaning of a reduced distance and characterizes the extent to which the nonlinearity of the system manifests itself. When the angular coefficient is small, the nonlinear properties manifest themselves weakly. The nonlinear properties become appreciable when Z is comparable with unity. Finally, they manifest themselves strongly when Z is greater than unity. The reduced distance depends on the true distance nonlinearly. As $x \rightarrow \infty$, the reduced distance tends toward its limiting value Z_∞ :

$$Z_\infty = \frac{\alpha \omega U_0}{\delta}, \quad (20)$$

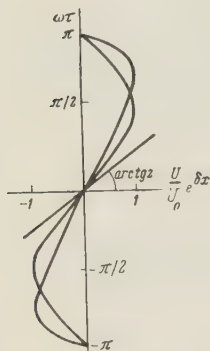


Figure 2

which can be smaller than unity in the case of sufficiently large attenuation. This means that, when the attenuation is sufficiently large, the distortion of the wave is small at any distance from the input of the system and that it only decays in amplitude (Figure 3a).

Solution (18) has fundamentally different forms when the reduced distance (19) is greater than or smaller than unity. When Z is greater than unity, then, as seen from Figure 3, c, d, the solution $U(\tau)$ is multivalued, which is a physical impossibility. This means that only a part of solution (18) describes the shape of the wave in the region $Z > 1$. Starting with a reduced distance equal to unity (Figure 3b) the wave becomes a shock wave, and the solution describing its form becomes discontinuous. The connection between the values of the current, the charge and the voltage in front of the shock wavefront and behind it is given by the equations

$$Lv_0(I_1 - I_2) = U_1 - U_2, \quad v_0(Q_1 - Q_2) = I_1 - I_2, \quad (21)$$

where v_0 is the velocity of the wavefront, while the subscripts 1 and 2 refer to the values before and after the wavefront. Equation (21) can be easily obtained by writing the condition for the continuity of the field at the discontinuities of the coordinate system moving with the velocity v_0 . If we eliminate the values of the discontinuities of the current from these equations and consider that $U_1 = U_2$ at the discontinuity, we obtain an equation for the velocity of the shock wavefront

$$v_0^2 = \frac{1}{L} \frac{U_2 - U_1}{Q_2 - Q_1} = \frac{1}{LC}, \quad (22)$$

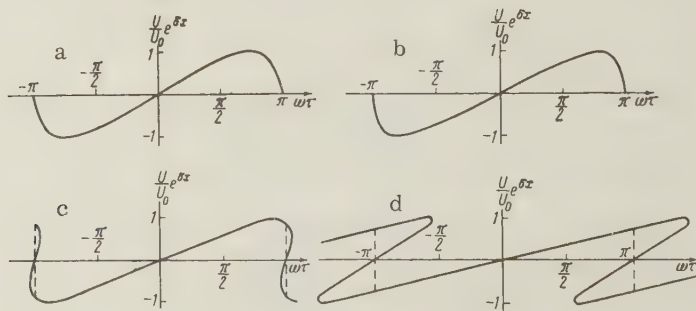


Figure 3

i.e., the wavefront moves with the velocity of propagation of small perturbations. In Figure 3c and d, the shock wavefront is shown by dotted lines.

It is necessary to consider especially the question of whether solutions in the form of a traveling wave can exist in the region of the discontinuous solutions of equations (1). This is connected with the fact that in a nonlinear system a voltage jump is equivalent to a jump in the values of the parameters in a linear system, and from this discontinuity waves should be reflected. Nonetheless, in the given case, as in the acoustic case of waves with sufficiently small amplitudes [6], no such reflection exists. This is the result of the fact that the values of the voltage and current in (21) before and after the discontinuity satisfy the equation (16) which is integrated in the neighborhood of the discontinuity and which is generally correct for any two points 1 and 2:

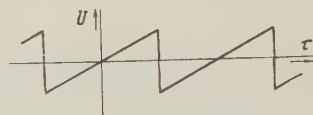


Figure 4

$$I_1 - I_2 = \sqrt{\frac{C}{L}} (U_1 - U_2) + \frac{D}{2\sqrt{LC}} (U_1^2 - U_2^2) + \frac{G}{2\sqrt{LC}} \int_{U_2}^{U_1} \frac{U dU}{\partial U / \partial t} \quad (23)$$

The two last terms in the right member (23) are equal to zero in the neighborhood of the discontinuity, and equation (23) transforms to the first of the equations (21). This proves the absence of reflection from the discontinuity.

From (20), we obtain the critical value for the input amplitude of the signal

$$U_{cr} = \frac{\delta}{\omega \alpha}, \quad (24)$$

below which the reduced distance can never be greater than unity. In this case, no shock wave is formed at any distance from the input.

From Figure 3c and d, it is seen that, starting with the reduced distances $Z = \pi/2$, the shape of the wave becomes almost saw-toothed (Figure 4). This conclusion is general, independently of the shape of the input voltage, if only it is periodic. At sufficiently large reduced distances, the shape of the wave becomes saw-toothed with decaying amplitude. When its saw-tooth voltage is applied to the input, the wave propagates along the line without any distortion of shape. This means that, in the case of a strongly dispersing line, there exists a stationary shape of the wave, which is propagated along the system without distortion. However, here there is an essential difference in the case of a dispersing line. This difference consists of the fact that the amplitude of this wave will always decrease even in the case where the parameters of the system are purely reactive.

Indeed, as can be seen from Figure 3d, the amplitude of the shock wave decreases with distance in the case $G = 0$ as

$$U \sim \frac{U_0}{1 + \alpha \omega U_0^2 x}. \quad (25)$$

The cause of its decrease is the specific attenuation of shock waves (cf. the acoustic case [6]). This specific attenuation takes place because, in accordance with the equation

$$v = \frac{1}{\sqrt{L(C + 2DU)}}, \quad (26)$$

where v is the velocity of a point on the wave profile, the positive part of the half-period ($U > 0$) moves more slowly than the negative part ($U < 0$). As a result, there is interference between these two parts of the wave and they cancel one another.

In this connection, it is interesting to note that the form of the law of conservation of energy usually adopted is

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial S}{\partial x} + GU^2 = 0, \quad (27)$$

where \mathcal{E} is the density of the wave's energy, and S is the density of the energy flux.

$$\mathcal{E} = \int_0^Q U dQ + \frac{1}{2} LI^2, \quad S = IU, \quad (28)$$

are uneven in the region of shock waves. According to (25), in the case $G = 0$ the average flow of energy \bar{S} decreases with distance according to the equation

$$\bar{S} \sim \frac{1}{2} \sqrt{\frac{C}{L}} \frac{U_0^2}{(1 + \alpha \omega U_0 x)^2}, \quad (29)$$

while according to the Equation (27) it should remain constant. This abnormal phenomenon is explained by the influence of the series resistance R (Figure 1) on the wave process, even in the case where it is a vanishingly small magnitude.

3. INFLUENCE OF THE SERIES RESISTANCE

For simplicity, let us confine ourselves to the case where there is no leakage, G and the energy losses in the system are caused only by the series resistance R . The simplified equation corresponding to (11) for this case is

$$\frac{\partial U}{\partial x} + \alpha U \frac{\partial U}{\partial \tau} = \delta \frac{\partial^2 U}{\partial \tau^2}. \quad (30)$$

Here α has a previous meaning (12), while, in distinction from (12), δ is equal to

$$\delta = \frac{1}{2} RC \sqrt{LC}. \quad (31)$$

The substitution [7, 8]

$$U = -\frac{2\delta}{\alpha} \frac{\partial v}{\partial \tau} \quad (32)$$

reduces Equation (30) to a linear equation of the same type as the heat equation

$$\frac{\partial v}{\partial x} = \delta \frac{\partial^2 v}{\partial \tau^2}, \quad (33)$$

For this case, when there is a sinusoidal voltage at the input (17), the boundary condition for (33) has the form

$$\text{and } x = 0 \quad v = \frac{\alpha U_0}{e^{2\omega\delta}} \cos \omega\tau. \quad (34)$$

The solution of Equation (33) with boundary condition (34) can be written in the form of an integral or a Fourier series:

$$v = \frac{1}{2\sqrt{\pi\delta x}} \int_{-\infty}^{\infty} e^{\frac{\alpha U_0}{2\omega\delta} \cos \omega y - \frac{(\tau-y)^2}{4\delta x}} dy = 2 \sum_{n=0}^{\infty} J_n \left(\frac{\alpha U_0}{2\omega\delta} \right) e^{-\omega^2 n^2 \delta x} \cos \omega n \tau. \quad (35)$$

Here J_n is Bessel's function for an imaginary argument. If the wave process is represented in the form of a spectrum of harmonic components, it will be found that the high frequency components are very strongly damped. In the most interesting case

$$R\omega C \ll \frac{DU_0}{C} \quad (36)$$

this representation describes the wave process only at fairly large distances from the input, where

$$x \gg \frac{2}{\omega^2 \delta} \quad \text{or} \quad \gamma x \gg 4Q_0, \quad (37)$$

Here $Q_0 = 1/R\omega C$ is the quality factor of the system and $\gamma = \omega \sqrt{LC}$ is its wavenumber. In this case, the shape of the wave is sinusoidal in the first approximation:

$$U \simeq \frac{2\delta\omega}{\alpha} \frac{J_1 \left(\frac{\alpha U_0}{2\omega\delta} \right)}{J_0 \left(\frac{\alpha U_0}{2\omega\delta} \right)} e^{-\omega^2 \delta x} \sin \omega\tau \simeq \frac{1}{Q_0} \left(\frac{C}{DU_0} \right) U_0 e^{-\omega^2 \delta x} \sin \omega\tau. \quad (38)$$

All the higher components on the spectrum of the wave formed in this first portion of the line where

$$\alpha\omega U_0 x \lesssim 1 \quad \text{or} \quad \gamma x \lesssim \frac{C}{DU_0}, \quad (39)$$

has been so damped because of the uneven attenuation that their amplitudes are considerably smaller than the amplitudes of the main component (38) at distances defined by (37).

To analyze the process of formation and "spreading out" the wave which is almost a shock wave, the integral representation (35) of solution (33) should be used. If the integral is computed by the saddle-point method, the corresponding analytic expression [9] can be obtained. From these expressions, it follows that up to the value

$$\gamma x = \frac{C}{DU_0} \left(1 - \frac{1}{Q_0}\right) \quad (40)$$

the effect of the resistance R on the process of formation of the wave which is almost a shock wave can be neglected in the first approximation.

The minimum possible duration of the quasi-shock of a wave occurring at distances of the order

$$\gamma x \simeq \frac{C}{DU_0} \frac{\pi}{2}, \quad (41)$$

corresponds to the phase interval $[\omega\tau]$:

$$[\omega\tau] = 2 \frac{1}{Q_0} \left(\frac{C}{DU_0}\right). \quad (42)$$

In the further motion of the wave, its shape can be described by the equation

$$U = \frac{U_0}{1 + \alpha\omega U_0 x} \left[\omega\tau - \pi \operatorname{th} \frac{\omega\tau}{\Delta} \right], \quad -\pi \leq \omega\tau \leq \pi, \quad (43)$$

where Δ gives the duration of the quasi-shock.

$$\Delta = \frac{1}{\pi} \frac{1}{Q_0} \frac{C}{DU_0} \left[1 + \frac{DU_0}{C} \gamma x \right]. \quad (44)$$

For values of γx of the order of $4Q_0$, the duration of the "front" is of the order of $\pi/2$ and the wave acquires in effect a practically sinusoidal shape.

Let us now consider the question of the flow of energy in the system. In the given case where $G = 0$ and $R \neq 0$, the equation corresponding to (27) has the form

$$\frac{\partial \xi}{\partial t} + \frac{\partial S}{\partial z} = RI \frac{\partial^2 I}{\partial z^2}, \quad (45)$$

Here, in the expressions for the energy density ξ and the energy flux density S , the voltage U refers to the voltage at the capacitance component U_C (Figure 1). Thus,

$$U = U_C + R \frac{\partial Q}{\partial t}. \quad (46)$$

It follows from (43) that in a region where a quasi-shock has formed, the right member of (45) is equal to zero everywhere during the period of τ , except for the small region of the quasi-shock, whose width is equal to Δ . As R decreases, this region becomes narrower, while the value of the right member of (45) increases proportionally to $1/R$ in this region. Therefore, when the average is taken over the period,

$$\frac{\partial \bar{S}}{\partial z} \neq 0$$

even in the case where $R \rightarrow 0$. This means that the specific attenuation of shock waves is caused by the resistance R , which is connected in series with the capacitive element, and that this resistance plays an essential role even when it is vanishingly small.

CONCLUSION

When radio waves are propagated along a weakly absorbing line with nonlinear distributed parameters, the shape of the wave is distorted. The wave becomes a shock wave and is

gradually transformed to a saw-tooth wave independently of the wave of the periodic voltage supplied. In the case where the part of the losses increasing with frequency is neglected (i.e., the distributed resistance in parallel to the capacitance is ignored), the spatial scales of this transformation are determined by the reduced distance (equation 19). As the input amplitude decreases, the distance at which a discontinuity in the profile of the wave begins to form is increased and becomes infinite at a certain finite value of the input amplitude.

When a shock wave is propagated along a line, specific scattering of the wave energy takes place. This is still true even when the parameters of the line are purely reactive. The attenuation of the wave is described by (19).

If the losses which increase in the frequency are taken into account, the steep front of the shock wave begins to "fall" gradually and the wave becomes sinusoidal.

The maximum number of the harmonic component which can be obtained in a system is determined from (42) by multiplying the quality factor of the system by the nonlinear parameter (the modulation factor of the capacitance).

The author thanks Professor G. Hefner (Stanford University, USA) for his interesting discussion on the topic of this article.

REFERENCES

1. L. K. Zaremko, V. A. Krasil'nikov, Problems of Propagation of Ultrasonic Waves of Finite Amplitude In Liquids, 1959, 68, 4, 687.
2. A. V. Goponov, G. I. Freidman, On the Theory of EM Shock Waves in Nonlinear Media, ZhETF, 1959, 36, 3, 957.
3. A. V. Goponov, G. I. Freidman, On the Theory of EM Shock Waves in Nonlinear Media, Izv. Vuzov, MVO SSSR (Radiofizika), 1960, 3, 1, 79.
4. R. Landauer, Parametric amplification along nonlinear transmission lines, J. Appl. Phys., 1960, 31, 3, 478.
5. G. M. Roe, M. R. Boyd, Parametric energy conversion in distributed systems, Proc. I. R. E., 1959, 47, 7, 1213.
6. L. D. Landau, Ye. M. Lifshits, Mechanics of Continuous Media, Izd. 2-ye GITTL, 1953, pp. 457-462.
7. E. Hopf, The partial differential equation $U_t + UU_x = \mu U_{xx}$, Commun. Pure and Appl. Math., 1950, 3, 3, 201.
8. J. D. Cole, Quasi-linear parabolic equation occurring in aerodynamic, Quart. Appl. Math., 1951, 9, 3, 225.
9. S. I. Soluyan, R. V. Khokhlov, On the Propagation of Finite-Amplitude Acoustic Waves in a Dissipating Medium, Vestn. Mosk. Univ., 1961, 3, 47.

Submitted to the Editors 24 August 1960

INDUCED DIPOLE IMPEDANCES IN A WAVEGUIDE

L. S. Benenson

The report derives design formulas for natural and induced impedances of current harmonics (the superposition of which may approximate any current distribution) distributed over a tubular dipole. The cases of placement of the dipole in a metal waveguide and in a

bidimensional rod grating are discussed. Design curves are presented which illustrate practical application of the formulas.

In the solution of a number of electrodynamic problems dealing with periodic structures consisting of dipoles it is necessary to find the impedance of the dipole (as a structural element) and also the current distribution at it. If the latter is presented as a superposition of the current harmonics forming a completely orthogonal system at the dipole surface, then both these problems reduce to calculation of the amplitudes of the individual current harmonics a_j and the impedances Z_{jk} which they induce in one another.

Amplitudes a_j may be determined from a system of linear equations in the induced-emf method and the radiation impedance Z is

$$Z = \sum_{j,k} a_j a_k^* Z_{jk}. \quad (1)$$

Since the coefficients of unknown amplitudes a_j in the mentioned equation are also Z_{jk} , it is immediately evident how important it is to know the values of Z_{jk} .

Below we present the calculation of Z_{jk} in the problem of a dipole located in a bidimensional grating or in a rectangular waveguide. Calculations are performed by a method based on the equivalence of a dipole within a waveguide and in a bidimensional grating obtained from the mirror image of a dipole within the walls of a waveguide. This permits extremely simple treatment of the thickness of the dipole and the nonuniformity of current distribution over the cross section. We will note that the mentioned equivalence was used in [1] in the solution of the problem of a dipole dividing a metal waveguide.

1. Let us consider the problem of a cylindrical rod dividing a rectangular waveguide with two metal and two magnetic (that is, ideally magnetoconductive) walls (Figure 1). If within the waveguide there is propagated a fundamental TEM wave, the current within the rod will have only one component (K_y) parallel to the axis of the rod, with distribution, constant along the rod, presentable in the form of a superposition of azimuthal harmonics:

$$K_y = \sum_{n=0}^{\infty} a_n \cos n\varphi, \quad (2)$$

where angle φ is read from the direction of the waveguide axis (z).

The field excited by the rod within the waveguide is equal to the field of a grating of cophasal rods obtained by the mirror image of a dipole within the waveguide walls, wherein distribution of currents in all the rods is expressed by formula (2). In order to determine the induced impedances, it is necessary to sum the fields of all these rods traversed by harmonics of one mode. The field of the n -th harmonic of one rod may be calculated by means of the Hertzian electric vector $\vec{\Pi}_{en} = \vec{\Pi}_{yn}^* \vec{i}_y$,

$$\Pi_{yn} = -\frac{\pi a Z_0}{2k} a_n J_n(ka) H_n^{(2)}(kR_0) \cos n\varphi. \quad (3)$$

Since for calculation of Z_{jk} it is necessary to find the field at the surface of the zeroth rod, in (3) it is convenient to transfer to another cylindrical system of coordinates y, r, ψ , the axis of which coincides with the axis of the zeroth rod (Figure 2). Then

$$R_0 = \sqrt{R_{|p|}^2 + r^2 - 2rR_{|p|} \cos \psi},$$

where $R_{|p|} = |p| L_x$ is equal to the distance between the old and new axes, that is, between the axis of the p -th and the axis of the zeroth rod; L_x is the period of the structure.

Applying the addition theorem for the Hankel function (see [2], formula (6.540.2)), we obtain for the expression

$$\Pi_{yn} = -\frac{\pi a Z_0}{2k} a_n J_n(ka) \sum_{m=-\infty}^{\infty} H_{n+m}^{(2)}(kR_{|p|}) J_m^*(kr) \cos m\psi. \quad (4)$$

Now let us add the fields of all the rods, that is, to the field of the zeroth rod [which we obtain from (3), assuming therein that $R_0 = r$] we add the sum for p from $p = \pm 1$ to $p = +\infty$ of

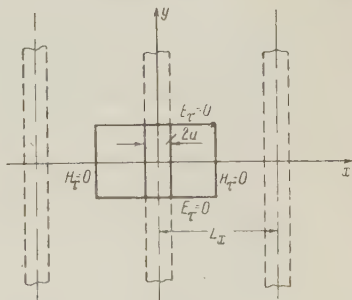


Figure 1

fields of the type in (4). In addition it is necessary to consider that if the origin of angles ψ and φ is taken to be the same for all rods (e.g., from the positive direction of the x-axis), then, for rods located to the left of the zeroth rod, it is necessary to replace ψ in (3) and (4) by $\pi - \psi$ and for rods to the right of the zeroth rod it is necessary to replace φ by $\pi - \varphi$. This will lead to the appearance of additional factors of $(-1)^n$ with $1 \leq p < \infty$ and of $(-1)^m$ with $-\infty < p \leq -1$. With this in mind we obtain for the Hertzian vector of the total field the expression

$$\Pi_{yn}^{(\Sigma)} = -\frac{\pi a Z_0}{2k} a_n J_n(ka) \left\{ H_n^{(2)}(kr) \cos n\varphi + \sum_{m=-\infty}^{\infty} J_m(kr) \cos m\varphi \times \right. \\ \left. \times \sum_{p=1}^{\infty} H_{n+m}^{(2)}(kpL_x) [(-1)^n + (-1)^m] \right\}. \quad (5)$$

The linear (per unit of length) impedance $Z_{n'n}$ induced by field $E_{yn} = k^2 \Pi_{yn}^{(\Sigma)}$ at the current harmonic of number n' of the zeroth rod is

$$Z_{n'n} = -\frac{1}{(2\pi a)^2} \int_0^{2\pi} \cos n'\varphi \frac{E_{yn}}{a_n} a d\varphi. \quad (6)$$

After simple calculation, we obtain

$$Z_{n'n} = \frac{kZ_0}{8} J_n(ka) \left\{ \epsilon_n \delta_{n'n} H_n^{(2)}(ka) + J_{n'}(ka) [1 + (-1)^{n+n'}] \times \right. \\ \left. \times \left[(-1)^{n'} \sum_{p=1}^{\infty} H_{n+n'}^{(2)}(kpL_x) + \sum_{p=1}^{\infty} H_{n-n'}^{(2)}(kpL_x) \right] \right\}, \quad (7)$$

where

$$\epsilon_n = \begin{cases} 1, & n \neq 0 \\ 2, & n = 0 \end{cases}, \quad \delta_{nn'} = \begin{cases} 0, & n' \neq n \\ 1, & n' = n \end{cases}. \quad (8)$$

With $n + n'$ equal to an odd number, $Z_{n'n} = 0$, that is, the even and odd harmonics do not interact.

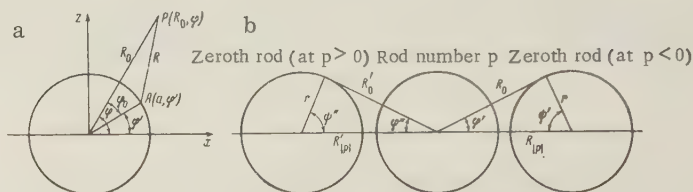


Figure 2

With $n' = n = 0$ (i.e., Z_{00}) expression (7) represents the radiation impedance of the rod in the zeroth approximation when it suffices to consider only the zeroth harmonic in the current distribution:

$$Z_{00} = \frac{kZ_0}{4} J_0(ka) \left\{ H_0^{(2)}(ka) + 2J_0(ka) \sum_{p=1}^{\infty} H_0^{(2)}(kpL_x) \right\}. \quad (9)$$

Upon the incidence of a TEM wave at a rod in the waveguide, the current harmonics are acted upon by the linear integral emf

$$f_0^{(n)} = \frac{1}{2\pi a} \int_0^{2\pi} \cos n\varphi E_y^{(TEM)} a d\varphi. \quad (10)$$

Since

$$E_y^{(TEM)} = e^{-ikz} = e^{-ika \cos \varphi} = \sum_{n=-\infty}^{\infty} e^{-\frac{i n \pi}{2}} J_n(ka) e^{in\varphi}, \quad (11)$$

for $f_0^{(n)}$ we find

$$f_0^{(n)} = i^{-n} J_n(ka). \quad (12)$$

Solving the system of equations of the method of induced emf's:

$$\sum_j a_j Z_{jk} = \frac{f_0^{(k)}}{2\pi a}, \quad (13)$$

we find, considering only the zeroth and first harmonics:

$$a_0 = \frac{f_0^{(0)}}{2\pi a Z_{00}}, \quad a_1 = \frac{f_0^{(1)}}{2\pi a Z_{11}}, \quad K_y = a_0 \left(1 + \frac{a_1}{a_0} \cos \varphi\right). \quad (14)$$

In order to evaluate the influence of the upper harmonics of current it is convenient to calculate the coefficient of reflection Γ from the rod in an equivalent waveguide.

Since interaction between the zeroth and first harmonics is absent, the total reflection coefficient is equal to the sum of the partial reflection coefficients resulting from both harmonics:

$$\Gamma = -\frac{R_{00}}{Z_{00}} - \frac{R_{11}}{Z_{11}}. \quad (15)$$

2. Great practical interest attaches to the problem of a dipole not wholly dividing a waveguide, for it corresponds to a metallodielectric in the form of a rod of finite length (Figure 3). In the rigorous solution of this problem it would be necessary to consider that at such rods, in addition to axial currents, there will also be azimuthal currents caused by higher modes excited in the equivalent waveguide by the rod itself (in this case, also, the TEM wave excites only axial currents). However, consideration of azimuthal currents would lead to extremely complex calculation. Since the role of these currents is not great, we shall disregard them in subsequent discussion. The distribution of axial currents at the surface of the rod will be represented in the form of a double Fourier series:

$$K_y = \sum_{p=0}^{\infty} \sum_{m=0}^{\infty} \frac{C_{mp}}{2\pi a} \cos \frac{(2m+1)\pi}{L} y \cos p\varphi, \quad (16)$$

$$-\frac{L}{2} \leq y \leq \frac{L}{2}, \quad 0 \leq \varphi \leq 2\pi$$

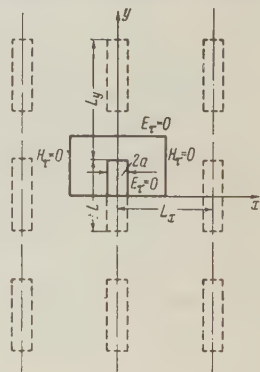


Figure 3

(L is the dipole length). The series in terms of coordinate y contains only odd cosine harmonics, since current distribution is symmetrical relative to the center of the dipole and at its ends vanishes (thus, strictly speaking, the investigated dipole represents a cylindrical tube rather than a solid rod). In calculating the field created by the dipole within the waveguide we proceed, as in the preceding case, from a single rod to a bidimensional grating obtained as the result of the mirror image of the dipole within the walls. The expression

$$\Pi_y = \frac{Z_0}{4\pi i k} \int_{s_0} K_y \frac{e^{-ikr}}{r} ds = \frac{Z_0}{4\pi i k} \int_{s_0} \sum_{n=-\infty}^{\infty} K_y \frac{e^{-ik\sqrt{R^2 + (y-y'-nL_y)^2}}}{\sqrt{R^2 + (y-y'-nL_y)^2}} ad\varphi dy', \quad (17)$$

$$-\frac{L_y}{2} \leq y \leq \frac{L_y}{2}$$

for the Hertzian vector of the field of one vertical row of rods is transformed by using the Poisson equation and the addition theorem (see [3], formula (14), and reference [2], formulas (6.541.2) and (6.540.2)) for the Hankel function. Then we obtain for rods of the zeroth row

$$\Pi_y^{(mp)} = -\frac{\pi a Z_0}{2k L_y} \sum_{n=-\infty}^{\infty} B_{nmp} J_p(\gamma_n a) H_p^{(2)}(\gamma_n R_0) \cos p\varphi e^{-i\frac{2n\pi}{L_y} y}, \quad (18)$$

for rods of the t -th row

$$\Pi_y^{(mp)} = -\frac{\pi a Z_0}{2kL_y} \sum_{n,l=-\infty}^{\infty} B_{nmp} J_p(\gamma_n t) J_l(\gamma_n r) H_{p+l}^{(2)}(\gamma_n t L_x) \cos l\psi e^{-i\frac{2n\pi}{L_y}y}, \quad (19)$$

where

$$B_{nmp} = 2 \cos \frac{n\pi L}{L_y} \frac{(-1)^m \frac{2m+1}{L} \pi}{\left(\frac{2m+1}{L} \pi\right)^2 - \left(\frac{2n\pi}{L_y}\right)^2} C_{mp}. \quad (20)$$

We find the total field at the surface of the rods of the zeroth row by summing expression (19) for t from ± 1 to $\pm \infty$ and adding the field of the zeroth row (18), wherein in (18) and (19) it must be assumed that $R_0 = r = a$:

$$E_y^{(mp)} = -\frac{\pi a Z_0}{2kL_y} \sum_{n=-\infty}^{\infty} B_{nmp} \gamma_n^2 \left\{ J_p(\gamma_n a) H_p^{(2)}(\gamma_n a) \cos p\varphi + \right. \\ \left. + \sum_{l=1}^{\infty} \sum_{l=-\infty}^{\infty} J_p(\gamma_n a) J_l(\gamma_n a) \cos l\psi H_{p+l}^{(2)}(\gamma_n t L_x) [(-1)^p + (-1)^l] \right\} e^{-i\frac{2n\pi}{L_y}y} \quad (21)$$

The impedance $Z_{p'm',pm}$ induced by the harmonic with subscript pm on the harmonic with subscript $p'm'$ is defined by the expression

$$Z_{p'm',pm} = \int_{-L/2}^{L/2} \int_0^{2\pi} \frac{E_y^{(mp)}}{(2\pi a)^2 C_{mp}} \cos p'\varphi \cos \frac{(2m'+1)\pi}{L} y a d\varphi dy. \quad (22)$$

Inserting expression (21) in place of $E_y^{(mp)}$ with consideration of (20) and integrating, we obtain

$$Z_{p'm',pm} = \frac{Z_0}{2kL_y} \times \\ \times \sum_{n=-\infty}^{\infty} \frac{(-1)^{m+m'} \gamma_n^2 \cos^2 \frac{n\pi L}{L_y} \left(\frac{2m+1}{L} \pi\right) \left(\frac{2m'+1}{L} \pi\right)}{\left[\left(\frac{2m+1}{L} \pi\right)^2 - \left(\frac{2n\pi}{L_y}\right)^2\right] \left[\left(\frac{2m'+1}{L} \pi\right)^2 - \left(\frac{2n\pi}{L_y}\right)^2\right]} J_p(\gamma_n a) \times \\ \times \left\{ \varepsilon_{x0} \delta_{pp'} H_p^{(2)}(\gamma_n a) + J_{p'}(\gamma_n a) [1 + (-1)^{p+p'}] \sum_{l=1}^{\infty} [(-1)^{p'} H_{p+p'}^{(2)}(\gamma_n t L_x) + \right. \\ \left. + H_{p-p'}^{(2)}(\gamma_n t L_x)] \right\}. \quad (23)$$

Let us note that in (23) the terms corresponding to the TEM wave coincide with expression (7), which refers to a rod wholly dividing the waveguide. Great practical importance attaches to the quantity $Z_{00,00}$, which is equal to the radiation impedance of a rod in the equivalent waveguide on the assumption that the current distribution is represented by one zeroth harmonic of series (16). For it we have

$$Z_{00,00} = \frac{(kL)^2 Z_0}{\pi^2 (kL_y)} \left\{ J_0(ka) H_0^{(2)}(ka) + 2J_0^2(ka) \sum_{l=1}^{\infty} H_0^{(2)}(kl L_x) - \right. \\ - \frac{4i}{\pi} \sum_{n=1}^{\infty} \frac{\cos^2 \frac{n\pi L}{L_y}}{\left[1 - \left(\frac{2nL}{L_y}\right)^2\right]^2} \left(\frac{\delta_n}{k}\right)^2 \left[I_0(\delta_n a) K_0(\delta_n a) + \right. \\ \left. + 2I_0^2(\delta_n a) \sum_{l=1}^{\infty} K_0(\delta_n l L_x) \right] \left. \right\}. \quad (24)$$

In (24) the terms resulting from the principal TEM wave and higher modes are written out individually, wherein it is assumed that $\gamma_n = -i\delta_n$ and the relation $H_0^{(2)}(-iz) = \frac{2i}{\pi} K_0(z)$ is used.

Expression (24) may be used directly for more precise calculations of phase velocity in the structure for formula (30) in [4]. It is seen from (24) that the propagating wave (TEM) causes an inductive reactive component of self-impedance of the rod (since $H_0^{(2)}(ka) = J_0(ka) - iN_0(ka)$, while $N_0(ka) < 0$ with $ka \ll 1$) and the higher wave modes cause the capacitive component (since $K_0(\delta_{na}) > 0$).

The integral emf caused by the TEM wave and acting on the current harmonic is

$$F_{pm} = \int_{-L/2}^{L/2} \int_0^{2\pi} \frac{1}{2\pi a} e^{-ikz} \cos p\varphi \cos \frac{(2m+1)\pi}{L} y a d\varphi dy = \frac{2Li^p(-1)^{m+p} J_p(ka)}{(2m+1)\pi}. \quad (25)$$

3. The induced impedances in the case of a rod in a metal waveguide may easily be determined by the same method used above. It is only necessary to consider that, since the waveguide walls are metallic, with mirror reflection [6] within them we obtain vertical rows of dipoles with alternating currents (that is, currents distinguished by the factor $(-1)^t$, where t is the number of the vertical row). In this case, for example, the expressions for Z_{00} and $Z_{00,00}$ are obtained from the corresponding formulas (9) and (24) by inserting factors $(-1)^t$ under the summation signs according to subscript t . For calculation of the phase velocity in a linear grating of rods in a metal waveguide we may use Eq. (30) from [4], which was derived for a metallodielectric in the form of a tridimensional periodic structure but remains valid in the case under discussion. It is necessary only to consider that in a metal waveguide the principal wave is not a TEM wave but an H_{10} wave. In this connection the wavenumber k in Eq. (30) of [4] must be replaced by the propagation constant γ_{10} of the H_{10} wave.

4. The method of calculation of finite width of the rods was discussed above in the example of rods within a waveguide, which corresponds to a metallodielectric in which the wave is propagated along one of the Cartesian axes of the periodic grating of the structure. No difficulty is encountered in generalizing the derived results for the case of arbitrary direction of propagation of waves in a metallodielectric. In order to simplify the equations let us limit ourselves to the case where the current at the rod is described by one zeroth harmonic. If the direction of propagation lies in a plane xoz perpendicular to the rods, the currents in the vertical rows of rods forming a bidimensional transverse (to the z -axis) grating are distinguished by factor $e^{-ith_x L_x}$, where t is the number of the vertical row, h_x is the x -component of the propagation constant of waves in the structure, L_x is the x -axis period of the structure.

Expressions for Z_{00} and $Z_{00,00}$ are obtained from formulas (9) and (24) by inserting the factors $\cos(ph_x L_x)$ and $\cos(th_x L_x)$, respectively, under the summation sign according to subscript t .

In the general case of arbitrary direction of propagation, current distribution at the rods of the transverse grating has the form

$$K_y = e^{-ith_x L_x - inh_y L_y} \psi(y), \quad (26)$$

where the numbers t and n define the position of the rod in the grating, h_x and h_y are the components of the phase constant, and L_x and L_y are the periods of the grating along the x - and y -axes.

Let us write function $\psi(y)$ in the form

$$\psi(y) = \bar{\psi}(y) e^{-ih_y y}, \quad (27)$$

isolating the factor $e^{-ih_y y}$ caused by the presence of propagation along the y -axis; in addition, $\bar{\psi}(y)$ may be presented in the form of a double Fourier series (16).

It is not difficult to see that the expression for $Z_{00,00}$ has the form

$$Z_{00,00} = \frac{L^2 Z_0}{\pi^2 k L_y} \sum_{n=-\infty}^{\infty} \frac{\gamma_n^2 \cos^2 \frac{n\pi L}{L_y}}{\left[1 - \left(\frac{2nL}{L_y}\right)^2\right]^2} J_0(\gamma_n a) \left\{ H_0^{(2)}(\gamma_n a) + \right. \\ \left. + 2J_0(\gamma_n a) \sum_{t=1}^{\infty} \cos(th_x L_x) H_0^{(2)}(\gamma_n t L_x) \right\}, \quad (28)$$

where

$$\gamma_n = \sqrt{k^2 - \left(\frac{2n\pi}{L_y} + h_y\right)^2}. \quad (29)$$

5. Figure 4 (1, precise calculation; 2, approximate calculation; heavy line, single rod) and Figure 5 (1, precise calculation; 2, approximate calculation) present curves for the ratio X_{00}/R_{00} plotted from precise formula (9) and approximate formula (when the rods of the grating are replaced by axial linear currents and Z_{00} is calculated as the impedance induced by these currents on the linear current located from the axis of the zeroth rod by a distance equal to its radius). The figures correspond to two values of transverse period of the structure ($L_x/\lambda = 0.4$ and 0.1) and are plotted as a function of rod radius a/λ . For comparison, one of the Figures gives the plot of $X_{00}/R_{00} = -N_0(ka)/J_0(ka)$ corresponding to a single rod [see formula (9)]. The table lists data for the relative deviation in the precise and approximate values of $X_{00}/R_{00} = \xi$.

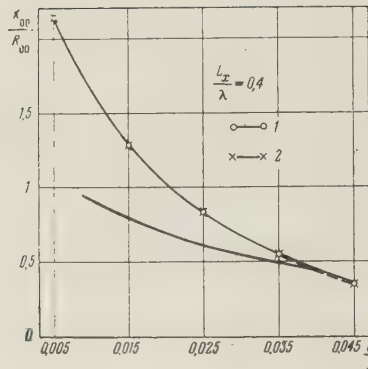


Figure 4

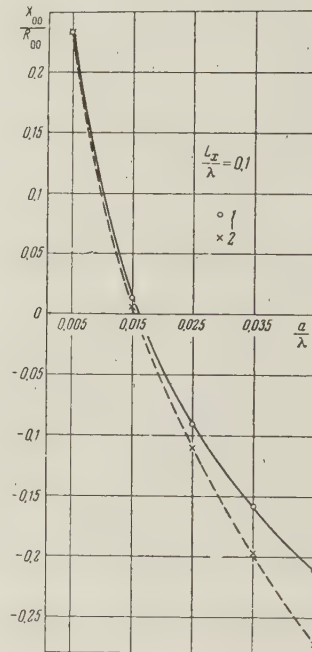


Figure 5

From Figures 4 and 5 and the table it is evident that the deviation between precise and approximate values of X_{00}/R_{00} increases with a decrease in period of the structure (when there is an increase in interaction of the rods of the grating) and with an increase in rod diameter, which was to be expected. At small periods ($L_x/\lambda = 0.1 - 0.2$) the corrections reach substantial value (on the order of 10-30%) even though the rods are not extremely thick ($a/\lambda \approx 0.025 - 0.045$).

$\frac{L_x}{\lambda}$	$\frac{a}{\lambda}$	$\frac{\Delta \xi}{\xi}, \%$
0.1	0.005	0.4
	0.025	22
	0.035	25
	0.045	29
0.2	0.005	0.04
	0.015	0.95
	0.025	8
	0.045	21
0.4	0.015	0
	0.035	1
	0.045	2

A comparison of the curves for X_{00}/R_{00} of a single rod and of a rod in a grating shows considerable difference between them, especially at small periods. In particular, at a single rod we always find $X_{00}/R_{00} > 0$, whereas, at a rod in a grating, X_{00}/R_{00} may be both positive and negative (the latter at small periods). It is possible, however, that in regions where $X_{00}/R_{00} < 0$ it is necessary to take into account the higher harmonics of current at the rods.*

*In calculating Z_{00} it is convenient to modify formula (9) by use of formulas (6.531.1), (6.532.3) and (6.534.3) in [2].

In the case of rods in a metal waveguide or in a structure with inclined incidence of the exciting wave the modification is achieved by means of formulas (6.532.1), (6.532.3), (6.533.3) and the

REFERENCES

1. L. Levin, *Modern Waveguide Theory*, transl. from English, IL, 1954.
2. I. M. Ryzhik, I. S. Gradshteyn, *Tables of Integrals, Sums, Series and Products*, GITTL, 1951.
3. Ya. N. Fel'd, L. S. Benenson, Calculation of phase velocities of waves in an artificial metallodielectric, *Radiotekhnika i Elektronika*, 1959, 4, 3, 417.
4. Ya. N. Fel'd, Incidence of electromagnetic waves at double infinite gratings, *Dokl. AN SSSR*, 1956, 107, 1, 71.

Submitted to the Editors 28 July 1960

MEASUREMENTS OF THE RADIAL ADMITTANCE OF AXIALLY SYMMETRICAL SYSTEMS WITH INDUCTIVE WALLS

L. V. Kasatkin

A method is proposed for measuring the radial admittance of axially symmetric systems with inductive walls. The essence of the method consists of experimental determination of the propagation characteristics of electromagnetic oscillations in a dielectric rod placed in the beam channel of the system under study. On the basis of the measured characteristics, the inductive susceptance at the edge of the dielectric rod can be determined from the theoretical equations derived here. The proposed method of measurement is in the frequency band beyond the critical frequency. The measured values of the admittance can be used to compute the amplification characteristics of electronic tubes in a drift tube with inductive walls under the condition that the amplification per unit length is small.

INTRODUCTION

The characteristics of the amplification of space-charge waves in electron beams in drift tubes with inductive walls outside the passbands depend essentially on the admittance of the walls reduced to the beam radius [1]. Thus, it can be shown that the range of amplification increases as the dependence of the inductive susceptance of the walls of the drift tube on frequency decreases; the amplification level increases as the absolute value of the inductive susceptance reduced to the beam radius decreases (to some limiting value), and so on.

Other conditions being equal, the admittance at the edge of the beam B_2 is determined by the magnitude of the inductive susceptance B_1 of the walls of the drift tube. Thus, in systems with axial symmetry [1], the inductive susceptance of the walls $Y_1 = -jB_1$ at the edge of a

ensuing formulas. In particular, with $ka \ll 1$, for R_{∞} we obtain the extremely simple expression

beam channel of radius r_1 is reduced to the edge of a beam of radius r_2 by the equation

$$Y_2 = -jB_2 = \frac{-j\omega\epsilon}{k_e} \left\{ \frac{I_1(k_e r_2) A(k_e r_1) + K_1(k_e r_2) B(k_e r_1)}{I_0(k_e r_2) A(k_e r_1) - K_0(k_e r_2) B(k_e r_1)} \right\}, \quad (1)$$

where

$$A = K_0(k_e r_1) + \frac{\omega\epsilon}{k_e B_1} K_1(k_e r_1); \quad (2)$$

$$B = I_0(k_e r_1) - \frac{\omega\epsilon}{k_e B_1} I_1(k_e r_1). \quad (3)$$

Here $I_0(k_e r)$, $K_0(k_e r)$ and $I_1(k_e r)$, $K_1(k_e r)$ are modified Bessel functions of zero and first order, respectively, $k_e = 2\pi/\lambda B_e$; $B_e = v_e/c$; and v_e is the velocity of the electron in the beam. Thus, it is possible to make a relative evaluation of the parameters of systems with inductive walls having a beam channel of definite diameter from the equations giving the inductive susceptance as a function of frequency $B_1(f)$.

Below we shall consider an experimental method of determining $B_1(f)$ from the propagation characteristics of electromagnetic oscillations in a dielectric rod with a dielectric constant $\epsilon = \epsilon_0 \epsilon'$, placed in the beam channel of the system under study. The essence of the method lies in the fact that the propagation characteristics in the dielectric are determined by the boundary conditions at the surface of the rod, i.e., they depend on the magnitude of the inductive susceptance of the drift tube.

1. MEASUREMENT OF THE RADIAL ADMITTANCES OF "SMOOTH" SYSTEMS WITH INDUCTIVE WALLS

To start with, let us consider the case of a "smooth" drift tube, in which the inductive susceptance $Y_1 = -H_\phi/E_z$ for $r = r_1$ remains constant along the entire drift. The wave equation for $r \leq r_1$ has the form

$$\nabla_{r,\phi}^2 E_z + \frac{\partial^2 E_z}{\partial z^2} + k^2 E_z = 0, \quad (4)$$

where $k = \frac{2\pi}{\lambda} \sqrt{\epsilon'}$. In the case of axial symmetry of the field, the solutions of this equation for a wave in the TM_{01} are given by the equations

$$E_z = A J_0(r \sqrt{k^2 - \gamma^2}) e^{j(\omega t - \gamma z)}, \quad (5)$$

$$E_r = A \frac{-j\gamma \sqrt{k^2 - \gamma^2}}{\gamma^2 - k^2} J_1(r \sqrt{k^2 - \gamma^2}) e^{j(\omega t - \gamma z)}, \quad (6)$$

$$H_\phi = A \frac{-j\omega\epsilon \sqrt{k^2 - \gamma^2}}{\gamma^2 - k^2} J_1(r \sqrt{k^2 - \gamma^2}) e^{j(\omega t - \gamma z)}. \quad (7)$$

$J_0(x)$ and $J_1(x)$ are Bessel functions of zero and first order, respectively. The susceptance of the boundary of the dielectric rod is

$$Y_1 = -\frac{H_\phi(r_1)}{E_z(r_1)} = \frac{-j\omega\epsilon}{k^2 - \gamma^2} \sqrt{k^2 - \gamma^2} \frac{J_1(r_1 \sqrt{k^2 - \gamma^2})}{J_0(r_1 \sqrt{k^2 - \gamma^2})}. \quad (8)$$

Let us represent the dispersion equation, which gives the phase velocity of propagation of electromagnetic oscillations in a dielectric, in the form of the condition for matched susceptances at the boundary (at $r = r_1$):

$$B_1 = \frac{kr_1 \sqrt{\epsilon'}}{120\pi} \frac{1}{\kappa} \frac{J_1(\kappa)}{J_0(\kappa)}, \quad \kappa = r_1 \sqrt{k^2 - \gamma^2}. \quad (9)$$

The function $\frac{B_1 \cdot 120\pi}{kr_1 \sqrt{\epsilon'}} = \frac{1}{\kappa} \frac{J_1(\kappa)}{J_0(\kappa)} = f(\kappa)$ is shown in Figure 1. The dotted curve corresponds to the conditions for retardation of the waves in the rod (phase velocity $v_\phi \leq c/\sqrt{\epsilon'}$), when $\kappa = j\bar{\kappa} = j\bar{r}_1 \sqrt{\gamma^2 - k^2}$, and the dispersion equation becomes

$$B_1 = \frac{kr_1 \sqrt{\epsilon'}}{120\pi} \frac{1}{\kappa} \frac{I_1(\bar{\kappa})}{I_0(\bar{\kappa})}. \quad (10)$$

The solid curve in Figure 1 corresponds to the condition $v_\phi \geq c/\sqrt{\epsilon'}$. For each fixed value of B_1 and for $v_\phi > c/\sqrt{\epsilon'}$, waves with different radial variations can be propagated. In the region of retardation of the electromagnetic oscillations ($v_\phi < c/\sqrt{\epsilon'}$), the dispersion equation (10) determines a single-valued relation between B_1 and κ .

As $B_1 \rightarrow \infty$, when the drift tube has ideally conducting walls, Eq. (9) transforms to the well-known equation giving the wavelength in a waveguide loaded with dielectric

$$\Lambda_g = \frac{\lambda}{\sqrt{\epsilon'}} \left[1 - \left(\frac{\lambda}{\sqrt{\epsilon'} \lambda_{cr}} \right)^2 \right]^{-1/2},$$

where $\lambda_{cr} = 2.61 r_1$.

When the value of the inductive susceptance B_1 is finite, the critical wavelength is determined from the condition $\gamma \rightarrow 0$ by means of the equation

$$B_1 \cdot 120\pi = \sqrt{\epsilon'} \frac{J_1(k_{cr} r_1)}{J_0(k_{cr} r_1)}. \quad (11)$$

In Figure 2, a family of curves of the function $B_1 \cdot 120\pi = F\left(\epsilon', \frac{\lambda_{cr}}{r_1}\right) = \sqrt{\epsilon'} \frac{J_1\left(2\pi \sqrt{\epsilon'} \frac{r_1}{\lambda_{cr}}\right)}{J_0\left(2\pi \sqrt{\epsilon'} \frac{r_1}{\lambda_{cr}}\right)}$ is plotted for different values of ϵ' . Accordingly, one of the methods of determining the inductive susceptance of the walls of the system consists of measuring the critical wavelengths when a dielectric rod of radius r_1 with a known ϵ' is placed in the drift tunnel. By placing rods with different values of ϵ' in the system and measuring the long wave boundary of the passband, it is possible to determine the frequency function $B_1(f)$ in accordance with Equation (11) and Figure 2.

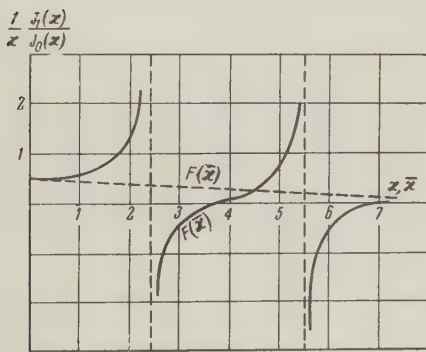


Figure 1. Variation in the function $\frac{1}{\kappa} \frac{J_1(\kappa)}{J_0(\kappa)}$ with κ for $v_\phi > c/\sqrt{\epsilon'}$ (solid curve), and variation of the function $\frac{1}{\kappa} \frac{I_1(\kappa)}{I_0(\kappa)}$ with κ for $v_\phi < c/\sqrt{\epsilon'}$ (dotted curve). This determines the relation of the propagation factor in the system with the quantity $B_1 \cdot 120\pi / \kappa r_1 \sqrt{\epsilon'}$.

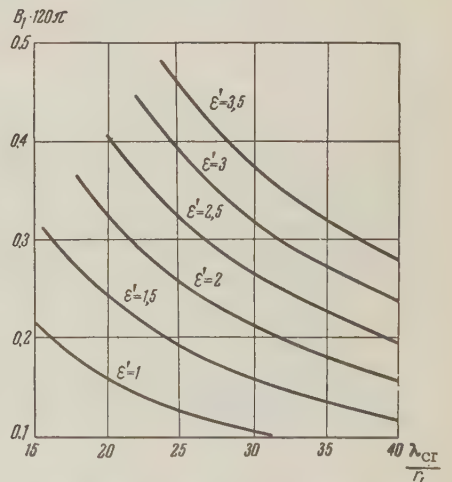


Figure 2. Family of curves of the radial inductive susceptance $B_1 \cdot 120\pi$ plotted as a function of λ_{cr}/r_1 for different values of ϵ' of a rod of radius $r = r_1$.

Another method of determining $B_1(f)$ consists of recording the dispersion function of a system with inductive walls with a dielectric rod inserted in the transit tunnel. In Figure 3, a family of curves of the function $\beta = v_\phi(c) = f(\lambda/r_1)$ constructed according to (9) and (10) is plotted for different values of the inductive susceptance $B_1 \cdot 120\pi$, with a permittivity of the rod $\epsilon' = 2.5$. When these curves are used with the measured data on the dispersion of the

system loaded with the dielectric, it is possible to determine the frequency characteristic of the inductive susceptance $B_1(f)$.

2. MEASUREMENT OF RADIAL SUSCEPTANCES OF PERIODIC STRUCTURES

When a structure with inductive walls is made in the form of a chain of connected cavities (Figure 4) for which the condition of axial symmetry at the boundary of the drift tunnel is not basically altered, it is necessary to add to the boundary conditions for $r = r_1$. In this case, it is necessary to take into account the space harmonics in the field of the structure when the fields are joined at the boundary of the rod in the same way as in the computation of the dispersion of a septate waveguide [2]. As known from [3], when the condition $D \ll \lambda\beta$ and $2t \ll D$ are satisfied, it is possible to consider only the fundamental space harmonics. The boundary conditions for $r = r_1$ have the form

$$E_{zI}(r_1) = \frac{1}{D} \int_{-\frac{d}{2}}^{+\frac{d}{2}} E_{zII}(r_1) e^{-j\gamma z} dz, \tag{12}$$

i.e., the intensity of the field in the dielectric rod is equal to the first space harmonic as a field in the structure:

$$\frac{1}{d} \int_{-\frac{d}{2}}^{+\frac{d}{2}} H_{\varphi I}(r_1) dz = H_{\varphi II}(r_1), \tag{13}$$

Figure 3. Family of curves of the delay factor $\beta = v\bar{c}/c$ plotted as a function of λ/r_1 , plotted for different radial susceptances $B_1 \cdot 120\pi$ at the boundary of the dielectric rod with $\epsilon' = 2.5$ on the radius $r = r_1$.

i.e., the intensity of the magnetic field in the structure is equal to the average value of the intensity in the dielectric rod in the interaction gaps. Here E_{zI} and $H_{\varphi I}$ are the components of the field in the dielectric rod for $r \leq r_1$; and E_{zII} and $H_{\varphi II}$ are the components of the field in the sections of the septate waveguide for $r \geq r_1$. The dispersion equation assumes the form

$$B_1 = \frac{M_e k r_1 \sqrt{\epsilon'}}{120\pi} \frac{1}{\kappa} \frac{J_1(\kappa)}{J_0(\kappa)}. \tag{14}$$

for septate waveguides with narrow radial chambers in which it can be assumed that the intensity of the electric field in the radial gap (at $r = r_1$) $E_z = \text{const} = E_0$ and that the phase shift in a chamber is $\psi < \pi/2$,

$$M_e = \frac{d}{D} \frac{\sin^2 \frac{\gamma d}{2}}{\left(\frac{\gamma d}{2}\right)^2}. \tag{15a}$$

For septate waveguides with thin-walled duct tunnels between the interaction gap in adjacent chambers, in which the field intensity in the gap at $r = r_1$ can be given by the equation

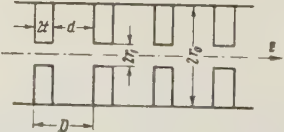


Figure 4. Chain of connected cavities.

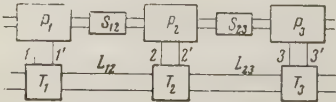


Figure 5. Equivalent circuit of a chain of connected cavities.

$$E_z = \frac{E_0}{\left\{1 - \left(\frac{2x}{d}\right)^2\right\}^{1/2}} [4], \quad (15b)$$

$$M_e = \frac{d}{D} \frac{\sin \frac{\gamma d}{2}}{\frac{\gamma d}{2}} J_0\left(\frac{\gamma d}{2}\right).$$

When a dielectric rod is placed in the drift tunnel of a system of coupled cavities, the radial inductive susceptance is no longer a single-valued function of frequency and depends on the phase velocity of the waves in the system. In the general case, this dependence is determined not only by the factor M [see (15a), (15b)], but also by the characteristic coupling elements between the cavities. Indeed, the equivalent circuit of a chain of cavities can be represented by Figure 5. Here, the six-poles P_1, P_2, P_3, \dots correspond to individual cavities coupled with one another by the quadripoles $S_{i, i+1}$ (for instance, slot couplings between the cavities) as well as by the line sections $L_{i, i+1}$, which are connected (through the transforming devices T_i) to the terminals $1-1', 2-2', \dots$ and which characterize the parameters of the dielectric rod in the drift tunnel of the system. In the special case where $\epsilon' = 1$ (there is no rod in the tunnel), the wave is not propagated along the chain of cavities in the region of frequencies outside the passband of the system and is rapidly damped as it moves away from the place of excitation. In this case, the admittance measured at the terminals $i, i+1$ at some fixed frequency ω_1 outside the passband $Y_1(\omega_1) = I_i/U_i$ (I_i and U_i are the current and voltage of the terminals of the i -th cavity) will be basically determined by the parameters of the i -th cavity and will depend only to a very small extent on the S couplings. When a dielectric rod is inserted in the drift tunnel, it is possible to select a value of its dielectric constant such that a propagating wave will appear in a system at the frequency ω_1 . When a system of this type is excited at the cavity terminals $1-1', 2-2', \dots$, voltages and currents will arise with a phase shift ψ relative to one another depending on the phase velocity of the waves v_Φ :

$$\psi = \frac{2\pi D}{\lambda} \frac{c}{v_\Phi}. \quad (16)$$

It is easily seen from the equivalent circuit that the radial admittance Y_1 in the traveling-wave operating mode depends on the coupling parameters S and on the phase shift between the cavities. Thus, the inductive susceptance measured when a rod with $\epsilon' > 1$ is inserted in the drift tunnel depends on v_Φ and may differ considerably from the admittance for $\epsilon' = 1$ in the absence of propagation. In this connection, it is necessary to find the value of β for which the measured quantity Y_1 corresponds to the system inductive susceptance which determines the characteristics of propagation of the space-charge waves when a beam passes through the drift space. We shall consider the case where an electron beam with a current density i_0 modulated at a frequency ω_1 lying outside the passband of the "cold" system is injected into the decelerating system. To simplify the problem, we shall assume that the charge density is constant throughout the cross section of this beam and that the beam totally fills the drift tunnel.* From [1], it is known that, when only the fundamental space harmonic is considered the electromagnetic field of the boundary of the drift tunnel in the system can be represented in the form

$$E_z = A J_0(T r_1) e^{j(\omega t - \gamma z)}, \quad (17)$$

$$H_\Phi = A \frac{-j\omega\epsilon_0}{\gamma^2 - k_0^2} T J_1(T r_1) e^{j(\omega t - \gamma z)}, \quad (18)$$

where

$$T^2 = (\gamma^2 - k_0^2) \left[\frac{\gamma_p^2}{(\gamma - k_e)^2} - 1 \right]. \quad (19)$$

By making the field continuous at the boundary of the drift tunnel, the following equation, which is similar to the dispersion equation (14); can be formed:

*When the beam diameter is smaller than the diameter of the drift tunnel, it is necessary to take into account equation (1) for transforming the impedances.

$$B_1 = - \frac{M_e k}{120 \pi} \frac{1}{(\gamma^2 - k_0^2) r_1} T r_1 \frac{J_1(T r_1)}{J_0(T r_1)}. \quad (20)$$

For definite values of the inductive susceptance B_1 , Equation (20) is satisfied for some complex value of the propagation factor

$$\gamma = \gamma_e \pm j \gamma_p q, \quad (21)$$

where γ_{pq} is the gain of electron waves in nepers per centimeter, and, $\gamma_p = \omega_p / v_e = \left[\frac{e}{m} i_0 / \epsilon_0 v_e^3 \right]^{1/2}$ is the wave number of the plasma.

Thus, in the given case, an increasing traveling wave which has a phase velocity determined by the accelerating potential (i.e., $v_\Phi = v_e$) and which is bound to the beam is propagated through the structure.

Suppose that the following conditions are satisfied at some fixed frequency ω_1 lying outside the passband of a structure: a) the phase velocity of the wave when there is a dielectric rod in the system is equal to the phase velocity of the amplified electron wave when a beam is injected in the tunnel, i.e., $v_\Phi = v_e$; b) the gain of the electron wave per unit drift length is negligible, i.e., $\gamma_{pq} \ll \gamma_e$. Then, the values of the inductive susceptances of the structure at the boundary of the drift tunnel [see Eq. (10) and (20)] can be regarded as identical in these two cases. Thus, it is necessary to find the radial inductive susceptance as a function of frequency for a given value at the same velocity, which is determined by the accelerating potential. For this purpose, a series of measurements of the dispersion function $\beta(\lambda)$ must be made in the system, with dielectric rods having different $\epsilon = \epsilon', \epsilon'', \epsilon''', \dots$ inserted in the drift tunnel (see Figure 6). For a given $\beta = \beta_e$, the wavelengths correspond to different dielectric constants ($\lambda_1, \lambda_2, \lambda_3, \dots$ in Figure 6) are determined from the dispersion characteristics. Then the values of $B_1 = f(\lambda/r_1, \epsilon')$ corresponding to the experimentally found quantities $\lambda_1 (\epsilon_1)$ are determined from Eq. (14), or from the function constructed according to (14) for the given value of β_e (see Figure 7). In Figure 7,* the dotted line represents the desired functions $B_1 = f(\lambda/r_1)$ for $\beta = \beta_e$.

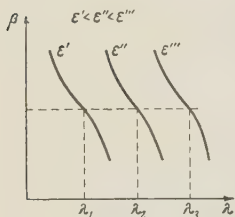


Figure 6. The dispersion function $\beta(\lambda)$ of the structure under study when dielectric rods with different values of $\epsilon = \epsilon', \epsilon'', \epsilon''', \dots$ are inserted in the drift tunnel.

In the case where the diameter of the dielectric rod is smaller than the diameter of the drift tunnel, the method described above makes it possible to determine the admittance of the system reduced to a given radius for each fixed value of $\beta_e = v_e/c$. Of course, a similar method of measurement can be developed for the study of systems with plane symmetry.

Let us note that the process of making measurements with different dielectric rods is necessary only when additional couplings (the S couplings in Figure 5) exist between the resonators. If these couplings do not exist (as in a separate waveguide with central drift spaces), B_1 as a function of the phase velocity will be entirely determined by the factor M_e . In this case, to determine $B_1 = f(\lambda, \beta)$, it is sufficient to measure the dispersion function with a single dielectric rod and then to carry out the corresponding recomputation for a given phase velocity $v_\Phi = v_e$ by means of Eqs. (14) and (15).

3. THE RADIAL ADMITTANCE OF A STRUCTURE WITH INDUCTIVE WALLS AS A FUNCTION OF FREQUENCY

The analysis we have given makes it possible to determine ways of making bandpass systems with inductive walls in which the frequency function $B_1(\omega)$ is small.

It is easily seen from Figures 6 and 7 that the curve showing the dependence of the measured inductive susceptance $B_1(\omega)$ on frequency for $\beta = \beta_e = \text{const}$ will be flatter (and the amplification in the drift phase will be wider), the greater the shift in wavelength of the dispersion function of the system when a dielectric rod is inserted in the drift tunnel. When the resonance method of measuring the dispersion function [5] is used, a shift in the dispersion curve in

*The functions in Figure 7 are constructed for the special cases where $\beta_e = 0.41$, $\epsilon_1 = 1$, $\epsilon_2 = 1.5$, $\epsilon_3 = 2$, $\epsilon_4 = 2.5$, $\epsilon_5 = 3$, $\epsilon_6 = 3.8$.

wavelengths corresponds to detuning of the resonator when the dielectric is inserted in it. If the resonator under study is represented in the form of two coupled "partial" regions corresponding to $r < r_1$ and $r > r_1$, the natural frequency of the resonator will be given by the equation

$$\omega_0^2 = \frac{\sum_{i=1,2} \frac{1}{\mu_i} \int_{V_i} (\text{rot } E_i)^2 dV_i}{\sum_{i=1,2} \epsilon_i \int_{V_i} E_i^2 dV_i}. \quad (22)$$

By the perturbation method, we find that when the drift tunnel ($r < r_1$), i.e., the first region, is loaded with a dielectric with $\epsilon = \epsilon_0 + \Delta\epsilon$, where $\Delta\epsilon$ is a small perturbation, the natural frequency of the resonator will be

$$\omega^2 = \omega_0^2 \left[1 - \frac{\Delta\epsilon}{\epsilon_{01}} \frac{W_{E_1}}{W_{E_1} + W_{E_2}} \right]. \quad (23)$$

Here W_{E_1} and W_{E_2} are the values of the stored electric energy in the first and second "partial" regions respectively.

The shift of the resonance frequency $\Delta\omega$ is given by the equation

$$\frac{\Delta\omega}{\omega_0} = - \frac{\Delta\epsilon}{2\epsilon_{01}} \frac{W_{E_1}}{W_{E_1} + W_{E_2}}. \quad (24)$$

Thus, the range of a system with inductive walls depends on the distribution of stored energy in the resonator and increases as the energy in the region of interaction increases relative to the total energy stored in the system.

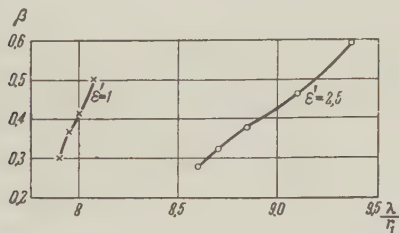


Figure 8. Dispersion function of a septate waveguide, $\beta(\lambda, r_1)$ without the dielectric rod ($\epsilon' = 1$) and with a dielectric rod with $\epsilon' = 2.5$ in the drift tunnel.

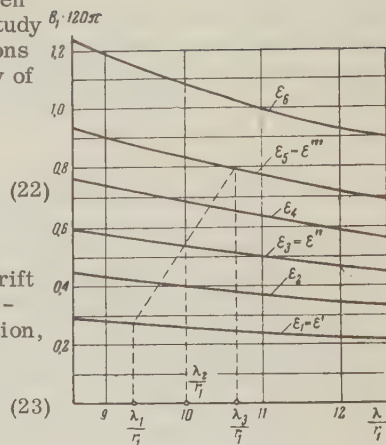


Figure 7. The family of curves $B_1 \cdot 120\pi = f(\lambda/r_1, \epsilon)$ constructed for a fixed value of β_0 and different values of the dielectric constants of the rod in the drift tunnel.

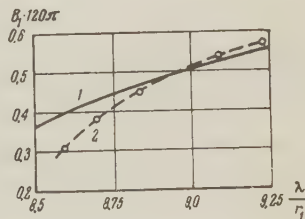


Figure 9. Theoretical (1) and experimental (2) curves of $B_1 \cdot 120\pi = f(\lambda/r_1)$.

4. EXPERIMENTAL VERIFICATION OF THE METHOD OF MEASUREMENT

Experimental verification of the method was carried out with a system with positive dispersion made in the form of a septate waveguide with central coupling spaces. At wavelengths greater than the long-wave limit of the passband of this waveguide, the radial admittance at the boundary of the drift tunnel was inductive and was given by the equation

$$Y_1 = -j \frac{1}{120\pi} \frac{[J_1(k_0 r_1) Y_0(k_0 r_0) - J_0(k_0 r_0) Y_1(k_0 r_1)]}{[J_0(k_0 r_1) Y_0(k_0 r_0) - J_0(k_0 r_0) Y_0(k_0 r_1)]}. \quad (25)$$

Figure 8 shows curves of the dispersion function of the septate waveguide without a dielectric rod ($\epsilon' = 1$) and with a dielectric rod with $\epsilon' = 2.5$ in the drift tunnel. Figure 9 contains curves of the function $B_1(\lambda/r_1)$: Curve 1, computed from Eq. (25); and Curve 2, determined from the measurements of the dispersion function in accordance with a graph of Figure 3 and on a basis of the coefficient M_{ϵ} . A comparison of these curves shows a fairly good agreement of the

theoretical and experimental results. The increase in the divergence in the region of shorter waves, with a phase shift in a chamber of $\psi > \pi/2$, appears to be connected with the fact that only the fundamental space harmonics have been considered. As in the computation for septate waveguides [3], it is possible to improve considerably the theoretical equations by considering two additional space harmonics in the system, (the first positive, and the second negative).

REFERENCES

1. S. K. Birdsall, J. R. Whinery, Waves in an electron stream with general admittance walls, *J. Appl. Phys.*, 1953, 24, 3, 314.
2. A. I. Akhizyer, Ya. B. Faynberg, Slow electromagnetic waves. *Uzbek. Phys. Nauk*, 1951, 10, 3, 321.
3. L. Levin, *Modern Waveguide Theory* (translated from English). 1954.
4. A. H. Beck, *Space-charge waves and slow electromagnetic waves*, Pergamon Press, 1958.
5. E. J. Nalos, Measurement of circuit impedance of periodically loaded structures by frequency perturbation, *Proc. I. R. E.*, 1954, 42, 10, 1508.

Submitted to the Editor 22 July, 1960

THEORETICAL AND EXPERIMENTAL INVESTIGATION OF FREQUENCY-DIVIDER KLYSTRONS WITH PRELIMINARY BUNCHING

O. A. Barmenkov, S. A. Kornilov, G. V. Lomakin

The report discusses theoretical investigation of frequency-divider klystrons with preliminary bunching, taking into account the influence of residual velocity modulation in the beam and nonlinear distribution of potential in the bunching space of the resonator. A quantitative comparison of theory and experiment in a reflex divider with preliminary bunching yielded satisfactory results.

INTRODUCTION

Investigation of klystron frequency dividers is of interest from the point of view of their application as the frequency converters for masers. Several works devoted to this subject have been published. References [1, 2] present a theory of double frequency division in klystrons of the transit-drift and reflex types. References [3, 4] and [2] present the results of experimental investigation of divider klystrons of various types showing qualitative agreement with theory.

For a quantitative comparison of experimental data obtained with a specially created divider klystron [4] it was necessary to take into account other theoretical factors, namely, the simultaneous effect of density and velocity modulation of the beam on the process of

frequency division and also the influence of nonlinearity of potential in the retarding space. The purpose of the work from which the data of this report were obtained was the carrying out of appropriate calculations and detailed experimental investigation of the divider klystron, permitting accurate comparison of theoretical and experimental data. Investigation was performed for the case of double frequency division.

1. THEORY OF DIVIDER KLYSTRONS WITH PRELIMINARY BUNCHING TAKING INTO ACCOUNT VELOCITY MODULATION AT THE INPUT OF THE RESONATOR

Dividers with a reflex and drift resonator are conveniently examined together since their descriptions are similar. These devices are represented in Figure 1.

The input sections of the two klystrons are identical and represent a modulating gap with drift sector Θ_1 , at the input of which there are alternating components of current and velocity, wherein the relation between these components depends on the space-charge density and length of the drift tube. We shall refer to the klystron in Figure 1a as a transit divider; the theory of this device, without consideration of the alternating component of velocity, was discussed in [1]. Correspondingly, we shall refer to the klystron in Figure 1b as a reflex divider. The theory of this device with $\Theta_1 = 0$ (i.e., considering velocity modulation only) was discussed in [2]. In subsequent discussion we shall assume that resonance of the input gap is absent (that is, the reflect divider employs scatter optics or any other device eliminating penetration of reflected electrons into the input gap).

Examination of these dividers is performed in much the same manner as in [1]. The only difference lies in calculation of the current induced at the grid of the output gap. Let us limit discussion to the case of small bunching of the beam reaching the input of the resonator of the klystron:

$$i_2 \simeq I_0 \left[1 + 2J_1 \left(X_{00} \frac{\sin h_1 l_1}{h_1 l_1} \right) \cos (2\omega t_2 - \Theta_{01}) \right],$$

$$v_2 \simeq v_0 \left[1 + \frac{X_{00} \cos h_1 l_1}{\Theta_{01}} \sin (2\omega t_2 - \Theta_{01}) \right].$$

As is subsequently explained, this does not impose serious limitations in comparing experiment and theory and at the same time does permit logical treatment of longitudinal debunching in the drift space Θ_1 .

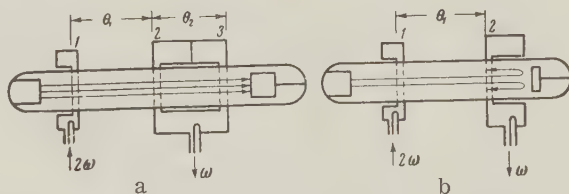


Figure 1. Transit (a) and reflex (b) dividers.

1) Modulating gap; 2 and 3) output resonator gaps.

Analyzing the convection current induced by the bunched stream at the output grids of the divider, for the principal harmonic of the divided frequency ω we obtain

$$i_3^{(\omega)} = A \sin \omega t_3 + B \cos \omega t_3,$$

where the Fourier coefficients are defined by the following integral:

$$B - jA = \frac{I_0 e^{-j\Theta_{02}}}{\pi} \int_{-\pi}^{+\pi} [1 + 2J_1(X_0) \cos(2\omega t_2 - \Theta_{01})] \times$$

$$\times e^{-j[\omega t_2 + X_v \sin(2\omega t_2 - \Theta_{01}) + X \sin(\omega t_2 - \varphi)]} d\omega t_2.$$

Here $X_0 = X_{00} \sin h_1 l_1 / h_1 l_1$ is the bunching parameter for bunching of the beam by the input signal, taking into account debunching in the drift space Θ_1 ; $X_v = \frac{\Theta_{02}}{\Theta_{01}} X_{00} \cos h_1 l_1$ is the bunching parameter in the resonator of the klystron, which bunching occurs due to residual velocity

modulation by the input signal at the input of this portion of the device. In the case of double signs (+ and -) the lower sign refers to the reflex divider. The integral is calculated by a Fourier-Bessel expansion:

$$B - jA = \pm 2I_0 e^{-j\theta_{02}} \sum_{p=-\infty}^{+\infty} (\pm 1)^p J_p(X_v) e^{-jp\theta_{01}} [J_{1-2p}(X) e^{-j(1-2p)\varphi} - \\ - J_{1+2p}(X) J_1(X_0) e^{-j\theta_{01}} e^{j(1+2p)\varphi} + J_{3-2p}(X) J_1(X_0) e^{j\theta_{01}} e^{-j(3-2p)\varphi}].$$

Limiting ourselves to values of X which are sufficiently small that we may neglect $J_3(X)$, $J_5(X)$, etc. in comparison with $J_1(X)$, we may obtain a result satisfactory for our purposes and maintaining with small X any degree of approximation required in practice. The assumption of small value of X is valid with small beam current near the threshold of self-excitation and also with large current near the maximum range of division, but, above all, it is valid at the edge of the division bandwidth where $X \rightarrow 0$ if only $\gamma [J_0^2(X_v) - m_e^2] \leq J_0(X_v)$, that is, in class-two resonant operation and in synchronization operation with sufficiently large input signal. This permits more accurate calculation of the division bandwidth — the chief characteristic of the frequency divider. After the mentioned simplification the expression for convention current in the output gap takes the form

$$i_3^{(\omega)} = \pm 2I_0 J_1(X) \{ [J_0(X_v) \sin(\theta_{02} + \varphi) \mp m_e \sin(\theta_{02} + \theta_{01} - \varphi)] \sin \omega t_3 + \\ + [J_0(X_v) \cos(\theta_{02} + \varphi) \mp m_e \cos(\theta_{02} + \theta_{01} - \varphi)] \cos \omega t_3 \},$$

where

$$m_e = J_1(X_v) \pm J_1(X_0) J_0(X_v) \pm J_1(X_0) J_2(X_v).$$

Inserting $i_3^{(\omega)}$ into the oscillation equations gives us

$$\frac{dX}{dt} = -\delta X - 2\delta\gamma J_1(X) [J_0(X_v) \mp m_e \cos(\theta_{01} - 2\varphi)], \\ \frac{d\varphi}{dt} = \frac{\omega^2 - \omega_0^2}{2\omega} \mp 2\delta\gamma \frac{J_1(X)}{X} m_e \sin(\theta_{01} - 2\varphi).$$

Here γ is the resonance parameter. For the transit divider $\gamma = k\beta_2\beta_3\theta_{02}I_0R/2U_0$, and for the reflex divider $\gamma = \beta_2^2\theta_{02}I_0R/2U_0$, where β_2 and β_3 are the coefficients of electron interaction for the second and third gaps of the klystron; k is feedback coefficient and is equal to the ratio of high-frequency at the second and third gaps.

The stationary solutions for both types of divider have the same form

$$\frac{2J_1(X)}{X} = \frac{J_0(X_v) \pm \sqrt{m_e^2 - [J_0^2(X_v) - m_e^2] \left(\frac{\omega^2 - \omega_0^2}{2\omega\delta} \right)^2}}{\gamma [J_0^2(X_v) - m_e^2]}.$$

A cursory examination of stability shows that the solution with a plus sign is always unstable and the solution with the minus sign has an extremely narrow region of instability at the extreme edge of the division range; consideration of the latter is of no practical importance.

From the last expression we obtain the relationships defining the relative division band-width $\Delta\omega/\omega$:

$$\frac{\Delta\omega}{\omega} = \frac{1}{Q} \frac{|m_e|}{\sqrt{J_0^2(X_v) - m_e^2}} \quad \text{for } \gamma [J_0^2(X_v) - m_e^2] \geq J_0(X_v), \\ \frac{\Delta\omega}{\omega} = \frac{1}{Q} \sqrt{\gamma^2 m_e^2 - [1 - \gamma J_0(X_v)]^2} \quad \text{for } \gamma [J_0^2(X_v) - m_e^2] \leq J_0(X_v).$$

The derived formulas indicate the different properties of transit and reflex dividers. Whereas in the transit divider modulation (both velocity and density modulation) of the beam at the input of the resonator of the klystron contributes to increasing $|m_e|$ (i.e., increasing the division bandwidth), in the reflex divider these phenomena have the opposite effect and with a certain relationship of parameters of the transit and reflex portions may even lead to disappearance of the effect of division.

Thus, an effort must be made to achieve predominance of one of the control mechanisms in the reflex divider (that is, to do this either with $l_1 \approx 0$ or with $l_1 \approx \pi/2h_1$). Choice of one or the other variant will be determined, on the one hand, by structural considerations and, on the other hand, by the attainable bunching angles in the transit and reflex portions. In particular, as will be seen below, solution of this problem may be substantially affected by nonlinearity of potential distribution in the retarding space.

As for the transit klystron, the combined action of bunching and velocity modulation of the beam contributes to increasing the division bandwidth, particularly in the case where the space charge is small and, despite a large angle Θ_1 , at the output of the drift space considerable residual velocity modulation is maintained.

2. INFLUENCE OF NONLINEARITY OF POTENTIAL DISTRIBUTION IN BUNCHING SPACE

Distribution of potential in the retarding space of a reflex klystron is usually characterized by a certain nonlinearity. In the reflex divider, in particular, this nonlinearity may be considerable, due to the necessity of creating scattering optics.

The influence of nonlinearity of potential distribution in the bunching space is considered in the first approximation in klystron theory by replacing transit angle Θ_0 in calculating the bunching parameter with bunching angle $a\Theta_0$, where [5]

$$a = -\frac{2U_0}{\Theta_0} \left(\frac{d\Theta}{dU_0} \right).$$

It is not difficult to confirm that consideration of nonlinearity of the retarding field in the case of a reflex divider in this same approximation also reduces to introduction of the bunching angle $\Theta_{02\text{bu}} = a\Theta_{02}$, which must be inserted instead of Θ_{02} into the formulas for X , X_V and γ . This generalization also applies to the transit divider.

Since, in practice, it is rather easy to attain values of a , the bunching gain, considerably greater than unity, the nonlinear potential distribution may be used for decreasing input power of the divider (that is, for improving its transfer coefficient). In this case, it is required that the electron stream be velocity-modulated at the input of the resonator.

3. KLYSTRON DESIGN AND RESULTS OF SIMULATION OF REFLEX OPTICS

An experimental investigation was performed with a reflex divider, the arrangement of which is shown in Figure 2. The tube construction was glass-metal with external resonators. The electron gun provided control of the beam independent of the potential at the diaphragms by regulation of the potentials at its anode and focusing ring. Short beam length with relatively large diameter (7-8mm) permits operation with electrostatic focusing. The input gap, designed for operation in the 3000 Mc range was formed by two copper diaphragms with plane grids made by the electric-spark method. The output gap was separated from the input gap by the drift space, which at a frequency of 2700 Mc had a transit angle of 50 radians. The mesh apertures of the output diaphragms have different diameters, which, in combination with the special reflector design, permits eliminating the penetration of reflected electrons into the

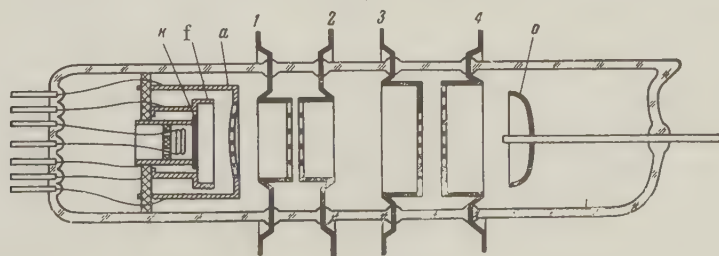


Figure 2. Experimental divider: k, f, a — cathode, focusing ring and anode of electron gun; 1, 2, 3, 4 — diaphragms; 0 — reflector.

input gap. Distortion of the regarding field, required for creation of scattering optics, is achieved by the introduction of a plunger.

Parameters of the reflector optics were quantitatively investigated by the analog method in an electrolytic tank. Space charge was disregarded. It was considered that electrons entering the retarding field have only an axial component of velocity.

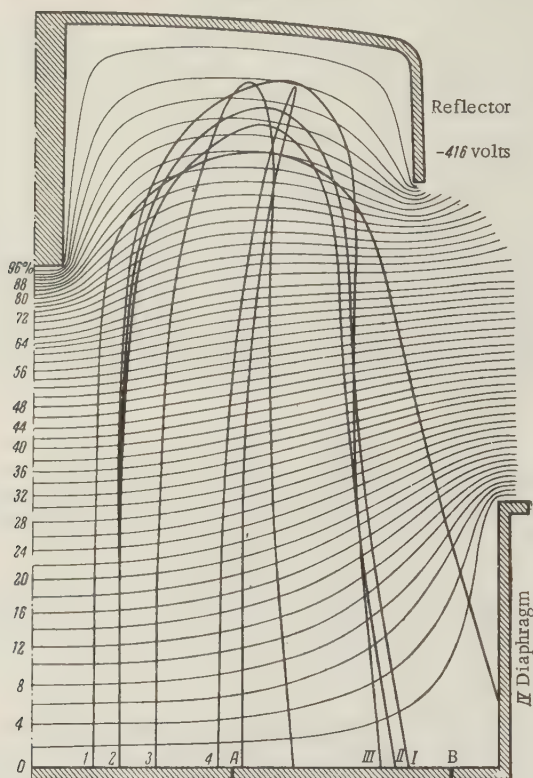


Figure 3. Retarding field and electron trajectories.

The equipotential lines (with successive 2% changes in supply voltage) and electron trajectories plotted therefrom by ordinary graphic methods are shown in Figure 3. Point A in this figure corresponds to the boundary of the mesh of the third diaphragm, point B corresponds to the boundary of the mesh of the fourth diaphragm. It is also seen from Figure 3 that with given potentials ($U_{ref} = -16$ v, $U_0 = 400$ v) the optics of the given reflector are satisfactory from the point of view of current passage. After reversal, electrons moving in trajectories 2, 3, 4, pass through the output gap of the resonator and are then intercepted by the solid portion of the third diaphragm. The electrons not entering the gap in reverse movement constitute approximately 15% of a total electron stream, if we consider that current density in the cross section of the beam is constant.

In order to determine the value of bunching gain a it is necessary to know the value of the derivative $(d\Theta_2/dU_E)_0$. For this purpose the trajectories 2-I, 2-II, 2-III ($U_0 = 400, 390, 410$ v, respectively) were plotted and by graphic integration the transit time of electrons along these trajectories was calculated. We then calculated the transit angles, obtaining a value of $\Theta_{02} = 31.3$ radians. From the plot of $\Theta_2(U_E)$ we determined the value of $(d\Theta_2/dU_E)_0 = 0.25$ r/v. Inserting into the expression for a the value of $(d\Theta_2/dU_E)_0$ and Θ_{02} , we obtain $a = 6.4$.

If we calculate the phase aberration index $\cos \psi$, where $\psi = \Theta_2 - 2\pi(n + 0.75)$ with $n = 4$ (the fourth zone of oscillation), we obtain $\cos \psi \approx 0.5$. Averaging was performed for three electrons moving in trajectories 2-I, 3, 4 corresponding to $U_0 = 400$ v.

Study of the static current flow confirmed the assumption adopted in analog simulation: under operating conditions the ray had extremely small convergence. In practice, its diameter upon entering the input gap was equal to the diameter of the aperture of the lower diaphragm at the output gap (that is, 7 mm), whereas at the input gap it did not exceed 8 mm.

High-frequency measurements indicated satisfactory operation of the scattering-retarding system as a whole. It always provided excitation of the output gap with faint traces of oscillation at the input gap.

On the basis of static measurements of current flow with positive potential at the reflector (which was greater than the diaphragm potentials in order to decrease errors associated with secondary emission) we determined the forward operating currents in the output gap and drift space which were required for calculation of a number of high-frequency parameters of the divider, particularly, longitudinal debunching in the presence of drift.

4. EXPERIMENTAL PROCEDURE

For a quantitative comparison of experimental data with theory it is necessary that the operating parameters of the divider be known with sufficient accuracy. These parameters include: voltages required at the gaps for determining the bunching parameters, the "hot" Q's of the resonators, the resonance coefficient of the output gap, and the bunching angles. Correct determination of these quantities is associated with certain difficulties due to nonlinear distribution of potential and phase aberrations in the retarding field.

Below we give a short description of the methods used by us in measuring these parameters.

Measurement of High-Frequency Voltage at Klystron Gaps. High frequency voltage was measured by means of detector probes which were calibrated by the method of total retarding of electrons.

For greatest accuracy the calibration point was chosen with a sufficiently large oscillatory voltage at the gap — of the order of 70 volts; then the detector current was plotted as a function of applied power, which permitted us to obtain a graduated curve down to fractions of a volt. In order to eliminate the influence of electron loading and resonance of the gaps, calibration was performed with extremely low beam current — less than 1 ma.

The coefficients of electron interaction in the gaps were calculated, which we were able to do with sufficient accuracy due to the planar construction of the gaps. With $U_0 = 400$ v they were calculated for an input gap $\beta_1 = 0.92$ (frequency 2700 Mc) and for an output gap $\beta_2 = 0.76$ (frequency 1350 Mc).

Determination of Resonator Parameters. (a) **Input Resonator.** The "cold" Q was measured from the resonance curve. Small changes in Q under the influence of electron loading were determined by means of a detector probe from the change in amplitude of oscillations in the resonator. In addition to the Q, it is of interest to determine the equivalent resistance of the resonator, calculation of which requires knowledge of the characteristic impedance ρ of the resonator. Calculation of the latter was difficult due to complexity of the geometry of a resonator with tuning screws. The quantity ρ was determined experimentally. In operation with matching of the resonator with the supply line and at low beam current (less than 1 ma) we obtain the following measurements: loaded Q, $Q_L = Q_0/2$, where Q_0 is the natural Q of the resonator; power of the incident wave P and voltage at the gap u. The measured quantities permit us to calculate ρ : $\rho = u^2/2PQ_0$. The resonator parameters so measured proved to be: $\rho = 32$ ohms, $Q_0 = 380$, $R = 12$ kilohms.

(b) **Output Resonator.** The output resonator has simple coaxial construction and, hence, its characteristic impedance was calculated. Measurement of the Q was complicated by the fact that it must be known in the operating state (that is, in the presence of a reflected electron stream). Its incidence at the peripheral solid portion of the lower diaphragm of the gap may give rise to additional electron loading, which is caused by the large current of slow secondary electrons. Hence, we resorted to the method of measuring the Q from the slope of electronic tuning in self-excited operation, taking into account all types of loss in the resonator. However, in the design formula it is necessary to take into account nonlinearity of potential distribution of the retarding field:

$$Q_n = -\frac{\omega_0}{2} \frac{\left(\frac{d\theta_2}{dU_{ref}} \right)_0}{\left(\frac{d\omega}{dU_{ref}} \right)_0}.$$

The value of the derivative $(d\theta_2/dU_{ref})_0$ in this formula is determined from the experimental curve for $\theta_2 = F(U_{ref})$, the nature of which depends upon the degree of nonuniformity of the retarding field.

Another method for determining the Q during oscillation conditions is that based on the slope of mechanical tuning of the klystron. Its advantage lies in independence of the results on the characteristics of the retarding field; however, in our case it provided lower accuracy due to imperfection of the tuning mechanism.

The measured values of the parameters of the output resonator were $\rho = 18$ ohms, $Q_0 = 400$, $R = 7.2$ kilohms.

Determining the Transit Angle in the Retarding Field. In our case, the most satisfactory of the known methods of measuring the transit angle in the retarding field of a reflex klystron is the method of variation of the natural frequency of the resonator, since it is not associated with a change in potentials of the reflector and resonator.

By means of this method, was obtained the transit angle in the retarding field as a function of the reflector and resonator potentials (Figure 4), which was required for a number of calculations.

Determining the Bunching Angle (Considering Nonlinearity of Potential Distribution in the Retarding Field). The bunching angle θ_{02bu} is determined from the bunching gain a : $\theta_{02bu} = a\theta_{02}$, where

$$a = -\frac{2U_0}{\theta_{02}} \left(\frac{d\theta_2}{dU_0} \right)_0 = -\frac{2U_0}{\theta_{02}} \left[\left(\frac{\partial\theta_2}{\partial U_{ref}} \right)_0 + \left(\frac{\partial\theta_2}{\partial U_0} \right)_0 \right]$$

Thus, knowing the experimentally derived relationship $\theta_2 = F(U_0, U_{ref})$, it is easy to calculate a . Our measurements indicated considerable nonlinearity: $a = 7.7$. This is in excellent agreement with the result obtained by simulation.

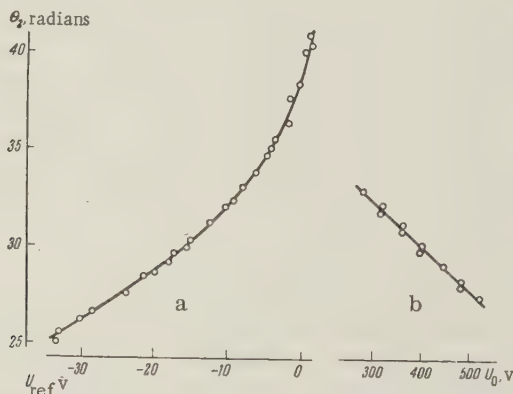


Figure 4. θ_2 as a function of U_{ref} with $U_0 = 400$ v (a) and of U_0 with $U_{ref} = -16$ v (b).

Determining the Resonance Parameter γ . The reflex optics of the investigated klystron possessed considerable phase aberration. Experimental determination of the aberration index as the ratio of the theoretical value of the threshold current in the output gap of the experimental value showed that $\cos \psi \geq 0.35$. This is in satisfactory agreement with the results obtained by simulation.

In order to avoid error due to phase aberration, the resonance parameter γ was not determined by calculation but from the ratio of the operating current in the gap to the threshold current (determined experimentally). With values of γ not differing extremely from unity the data of current flow permitted us to consider it to be proportional to the ratio of operating current of the cathode to its threshold value, which made determination of γ a simple operation.

Phase Measurements. In order to check frequency division and investigate the phase shift of input and output oscillation, depending on the operating conditions of the klystron, their frequencies were translated to several hundred kc and then the oscillations were compared on the screen of an oscillograph by means of Lissajous figures.

5. RESULTS OF EXPERIMENTAL INVESTIGATION

Below we present the results of experimental investigation of the described klystron in double-frequency division.

Figure 5 shows the dependence of the relative division bandwidth on the degree of bunching of the beam by input oscillation. The sum $X_0 + X_V$ is laid off on the abscissa axis wherein in calculating X_V the bunching gain in the retarding field was considered. It is of interest to note that in our case, despite the relatively large drift space $X_V/X_0 = 3.5$, that is, the influence of residual velocity modulation in the beam is noticeably predominant. If we disregard non-linearity of potential distribution in the retarding field, $X_V/X_0 = 0.45$ (that is, there is a significant difference in the result). In plotting the curves, we considered the longitudinal debunching in the drift space ($h_1 l_1 = 0.72$ with $\gamma = 1.6$) and the bunching gain in a nonuniform retarding field ($a = 7.7$). The transit angles in the drift space and retarding space were, respectively, $\Theta_{01} = 50$ radians, $\Theta_{02} = 30$ radians; the bunching angle in the retarding field $\Theta_{02bu} = a$, $\Theta_{02} = 230$ radians. In addition to the experimental curves (dash-dot), the figure presents theoretical curves.

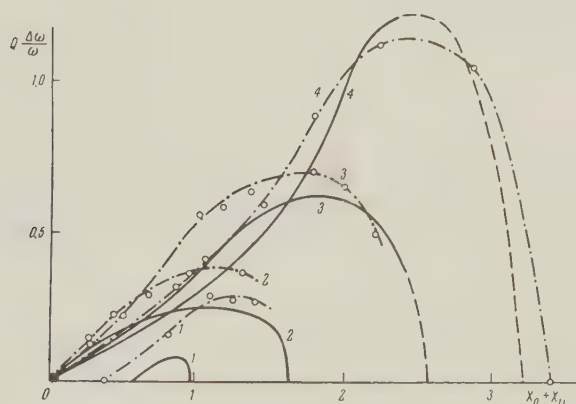


Figure 5. Relative division bandwidth as a function of the sum of parameters of bunching of the beam by the input signal:

- 1, $\gamma = 0.9$; 2, $\gamma = 1.0$; 3, $\gamma = 1.6$; 4, $\gamma \approx 3$; $U_0 = 400$ v; $V_{ref} = -16$ v.

Let us discuss the limits of applicability of the theoretical calculations. As stated above, our analysis assumes small values of X_0 (quasi-linear bunching operation in the drift space). Strictly speaking, this may be done as long as X_0 is small in comparison with unity. However, precise indication of the limit of values of X_0 at which the theoretical formulas become ineffective is difficult, since the higher harmonics of current in the beam at the input of the resonator (and, consequently, the harmonics of velocity) have relatively little effect on the

process of synchronization. This is particularly evident from [1], wherein it was shown that the error in determining the induced current when neglecting the effect of higher harmonics in the control beam remains less than 18% even with $X_0 = 1$. Thus, there is reason to believe that the actual range of application of the formulas given in Section 1 will be wider. In Figure 5, parts of the theoretical curves corresponding to $X_0 > 0.5$ are given by a dashed line, indicating the conditional nature of the comparison.

Satisfactory coincidence is observed even in the region of large values of $X_0 + X_V$, where theory does not guarantee accuracy. This is evidently associated with the fact that in frequency division the upper harmonics of a stream bunched at the input of the tube actually have little effect on the process of synchronization. An exception is found in the curves for voltage-controlled operation ($\gamma = 0.9$), which diverge. This is probably due to inaccurate determination of γ . Actually, in voltage-controlled operation the quantity γ has considerably greater effect on the division characteristics than in synchronization operation. Hence, the requirements for accuracy in determining γ increase.

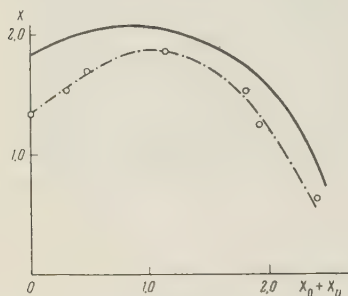


Figure 6. Amplitude characteristics of divider; $\gamma = 1.6$, $U_0 = 400$ v, $U_{ref} = -16$ v.

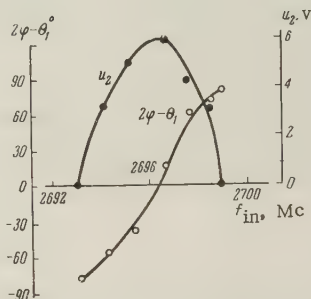


Figure 7. Resonance characteristics of divider: $U_0 = 440$ v; $U_{ref} = -13.5$ v; $\gamma = 1.6$; $u_1 = 7.7$ v; u_1 and u_2 , amplitudes of r-f voltage at the first and second gaps; f_{in} , frequency of input signal.

In order to confirm sufficient accuracy in determining the bunching parameters, we compared the theoretical and experimental values of the optimum bunching parameter with operation of the klystron in drift amplification. Departure did not exceed 15 percent. This check involves complex procedure, for determination of the bunching parameters is associated both with measurement of the gap voltages and with knowledge of the geometry of the beam and space-charge density.

Figure 6 presents the theoretical and experimental amplitude characteristics of the divider. The coincidence here is also satisfactory.

Finally, the resonance characteristics of division are shown in Figure 7. Here we plotted the amplitude of synchronized oscillations and change in their phase with a change in frequency at the divider input.

The present report has not dealt with multiple frequency division. Characteristics of the above klystron in frequency division with factors greater than 2, are presented in [4].

CONCLUSION

Satisfactory correspondence between the theoretical and experimental characteristics of a reflex divider with preliminary bunching has been obtained. It has been shown that the presence of the drift-space does not eliminate the need for considering residual velocity modulation within the beam, which has a particularly strong effect in the presence of increased bunching in the retarding field. The latter effect permits a substantial reduction in signal power at the divider input.

REFERENCES

1. S. A. Kornilov, Frequency division in a floating-drift klystron with three gaps, *Radiotekhnika i Elektronika*, 1958, 3, 4, 522.
2. Ye. N. Bazarov, M. Ye. Zhabotinskiy, Frequency conversion in a reflex klystron, *Radio-tekhnika i Elektronika*, 1959, 4, 2, 253.
3. Ye. N. Bazarov, M. Ye. Zhabotinskiy, Frequency division in a reflex klystron, *Radiotekhnika i Elektronika*, 1956, 1, 5, 680.
4. S. A. Kornilov, A. F. Yemelyanov, Experimental investigation of a klystron frequency divider with preliminary bunching, *Radiotekhnika i Elektronika*, 1960, 5, 2, 236.
5. D. Hamilton, J. Kuper, J. Knipp, *Klystrons and Microwave Triodes*, N. Y., 1948.

M. I. Kalinin, Leningrad
Polytechnic Institute

Submitted to the Editors
28 July 1960

NOTATION

- I_0, v_0 — direct components of convection current and electron velocity
 i_2, v_2 — alternating components of convection current and electron velocity at input of resonator of divider
 t_1, t_2 — moments corresponding to electron transit through first and second gaps
 t_3 — moment corresponding to electron transit through third gap for drift divider and reverse transit through second gap for reflex divider
 X_{00} — parameter for preliminary bunching of beam by input signal
 X — parameter for bunching of beam by oscillations of divided frequency
 ω — divided frequency
 ω_0 — natural frequency of output resonator
 φ — voltage phase of synchronized oscillations at second gap (input-voltage phase assumed to be zero)
 θ_1, θ_2 — transit angles in first and second bunching spaces, calculated for input and output frequencies, respectively
 θ_{01}, θ_{02} — transit angles of electrons with unmodulated velocity
 l_1 — length of first drift space
 h_1 — debunching parameter in drift space
 γ — resonance parameter
 δ, Q, R — attenuation, Q and equivalent resistance of output resonator
 U_0, U_{ref} — accelerating and retarding potentials
 U_e — electron energy

HARMONICS IN AN ELECTRON BEAM AT THE OUTPUT OF A TRAVELING-WAVE TUBE

V.I. Kanavets

Electron bunching is studied in the beam moving in a drift cylinder at the output of a traveling-wave tube at distances greater than the wavelength of the oscillations of the electron plasma. The bunching was studied on the basis of the third harmonic in the range of 10-cm wavelengths. It is shown that at these distances it is possible to obtain the third harmonic with the power of the same order of magnitude as the power of the harmonic in the traveling-wave tube. At large signal levels at the input of a traveling-wave tube, an increase of the amplitude of the third harmonic with distance, connected with the oscillations of the electron plasma in the flow, was discovered.

INTRODUCTION

At the present time, multicascade tubes having retarding structures and an electron beam using frequency multiplications have appeared. The electron buncher in these tubes is the traveling-wave tube. Thus, the frequency multiplier with two retarding structures [1, 2] is of this type. An advantage of this type of instrument is the large frequency-conversion coefficient (greater than unity) for a series of higher harmonics. Frequency multiplication with the use of electron bunching at the output of the traveling-wave tube is used in frequency-dividing devices [3], and can also be used in electronic parametric amplifiers with a low-frequency excitation [4]. Beams bunched in traveling-wave tubes are used to study different types of emission from electron bunches, such as Cherenkov radiations and radiation in inhomogeneities [5].

The devices which are described use harmonics of a beam which emerges from a traveling-wave tube and moves in space with the potential, which, in the majority of cases, is not equal to the potential of the retarding structure of the traveling-wave tube. It is possible to determine the main features of electron bunching in these devices by examining the electron bunching in a drift space at the output of the traveling-wave tube, with different potentials in this space. As far as we know, there is no theoretical and experimental work on this phenomenon in the literature. The published articles study electron bunching inside the retarding system of traveling-wave tubes. In our work on the nonlinear theory of traveling-wave tubes when the effect of the space charge is negligible [6], it is shown that the stream in these tubes has large harmonic components. The nonlinear properties of traveling-wave tubes taking into account the space charge are studied most completely in [7] and [8], but the beam harmonics are not considered in these works. The harmonics which appear when electrons are bunched in the drift space have been considered in the theory of klystron frequency multipliers for the case of initial velocity modulation of the beam [9]. Article [10] presents the nonlinear theory of electron bunching in two-cavity klystrons. The space charge forces are taken into account by means of a disc model of the electron beam. It is shown that the change in the amplitude of the harmonics along the beam has the form of a standing wave caused by the plasma oscillations in the beam. For a narrow beam, the wavelength of the standing wave depends on the number of the harmonic and decreases for higher harmonics.

The theory of the bunching of electrons in klystrons is generally not applicable to the description of the bunching in the flow at the output of traveling-wave tubes, since such a

stream contains initial current oscillations, at both the fundamental frequencies and at harmonic frequencies which are not taken into account in the theories. Modern traveling-wave tubes use electron beams with a current density of the order of 0.1 to 1 amp/cm² and more. The electric length of the traveling-wave tubes of the first cascade of multipliers, as well as of the following cascades, is equal to several times the wavelength of the plasma oscillations. Accordingly, an experimental study of the bunching of electrons in a beam which has emerged from a traveling-wave tube and moves in a drift space over distances much greater than the wavelength of the plasma oscillations, for various potentials of the drift space, is worth undertaking. The results obtained in this way can be used in examining the bunching in multistage tubes, for instance, in the helix of the traveling-wave tube in the second stage.

There are different methods for studying the amplitude of the alternating current components along a beam. Of these methods, the method using a movable probe is most applicable for the study of electron bunching. This method was used in a number of studies, for instance in [11], where a mobile resonator was used to study the waves in an electron beam. We used a similar method, but in distinction to work [11] we used a nonresonant wideband system to separate the energy of the harmonics.

1. EXPERIMENTAL TUBES AND MEASURING APPARATUS

Electron bunching in the beam at the output of a traveling-wave tube was studied by separating the third current harmonic with a movable probe. Fig. 1 shows the design of the experimental tube. The electron buncher is a traveling-wave tube ($\lambda = 30$ cm). The electron bunching begins under the action of the field of the helix of the traveling-wave tube and continues in the drift cylinder. The tube uses a hollow beam of electrons, produced by a ring

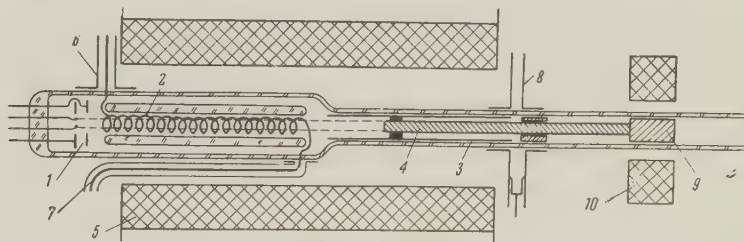


Fig. 1. Experimental tube: 1 — electron gun; 2 — helix of the traveling-wave tube; 3 — drift cylinder; 4 — moving probe; 5 — solenoid; 6 — input of the traveling-wave tube; 7 — output of the traveling-wave tube; 8 — output for the harmonic signal; 9 — cylinder of the moving device, made of a magnetic material; 10 — auxiliary solenoid.

cathode placed in the longitudinal field of the solenoid. The area of the emitting surface of the cathode is 2.2 mm²; the external diameter of the beam is 3.2 mm and the internal diameter 2.6 mm; the ratio of the diameter of the spiral to the average diameter of the helix to the average diameter of the beam is 2.2 ; the diameter of the helix is 8 mm; its pitch is 1 mm and its length 185 mm. The tube has a coaxial signal input and output. The electron beam goes from the output of the traveling-wave tube to a drift cylinder having an internal diameter of 8 mm. The distance between the helix and the drift cylinder (8 mm) is smaller than the length of the target waves at the fundamental frequencies. Inside the drift cylinder a circular probe 4.2 mm in diameter is moved, this probe being centered by means of a glass ring which is fitted about it. The distance l from the end of the helix to the front of the probe can be measured. This distance can be changed from $l = 0$ to $l = 26$ cm as the probe is moved.

The electron beam is focused inside the helix and the drift cylinder by the longitudinal magnetic field of a solenoid. The intensity of this field is 700 oersteds. The beam moves along the axis of the drift cylinder and is precipitated on the probe, which is an electron collector. A shortcoming of this method of separating the high-frequency energy is that the bunches in the beam emit a small amount of power (less than 1 μ w) in settling. However the use of this method in study of electron bunching is justified by the simplicity of the

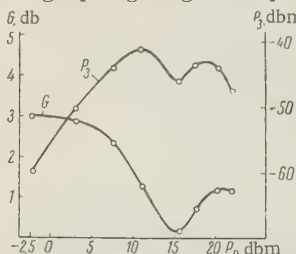
apparatus combined with the adequate sensitivity of the method. The probe on the drift cylinder forms a coaxial line, along which the radiated energy of the beam harmonic enters into the output device.

The output device is formed by a coaxial line — waveguide transition through glass. To supply energy from the coaxial line to the waveguide without large losses, quarter-wave coils are used. For correct removal of the energy of the harmonics, it is necessary that the coaxial line and the waveguide be carefully matched. If the matching is disturbed, a picture of the standing wave in the coaxial line will be recorded as the probe moves, changing considerably the picture of the harmonic signal. The matching is achieved by selecting the length of the coil and by adjusting the position of the plunger in the waveguide. The probe, which forms the internal conductor of the coaxial line, slides inside the coil cylinder, when it is moved, without however, disturbing the matching.

The moving device is formed of a steel cylinder which is attached to the probe and which moves in a glass tube under the action of an auxiliary solenoid. The signal of the third harmonic is fed from the output of the tube to a superheterodyne receiver and is indicated by a measuring instrument or an oscillograph screen. To take oscillograms, voltage from an AC circuit with a frequency of 50 cycles is fed simultaneously to the horizontal deflecting plate of the oscillograph and to the drift cylinder or the helix for modulating the constant voltages. The potential of the probe is always equal to the potential of the drift cylinder, and the potential of the drift cylinder can be changed within broad limits.

2. OPERATING CONDITIONS OF THE TRAVELING-WAVE TUBE

The measurements were conducted at a beam current $I = 3.5$ ma and with a potential of the helix $U_c = 500$ v. The wavelength of the oscillations supplied to the input of the traveling-wave tube was equal to 28.5 cm and the wavelength of the third harmonic to 9.5 cm. Figure 2 shows graphs giving the dependence of the electronic amplification of the traveling-wave tube G



and the power of the third harmonic signal at the output of the traveling-wave tube P on the power of the input signal P_0 . The graph for the power of the third harmonic was obtained with

Fig. 2. Dependence of the electronic amplification in the traveling-wave tube G and the power of the third harmonic signal in the plane of the end of the helix ($l = 0$) P on the signal power on the input of the traveling-wave tube P_0 .

the probe placed in the plane of the end of the helix, $l = 0$. The graph characterized the operating mode of the traveling-wave tube for different values of the power P_0 . For values of P_0 less than 2 mw, nearly linear operation occurred in the traveling-wave tube. The amplification differs insignificantly from the amplification of small signals, and the power of the third harmonic is much less than the maximum. For power P_0 greater than 2 mw, the operation is nonlinear. The amplification falls as P_0 decreases and reaches a first minimum at $P_0 = 36$ mw. The third harmonic signal in the plane of the end of the helix reaches a first maximum at $P_0 = 10$ mw. If the power of the signal at the input is greater than 10 mw, the amplification and the power of the third harmonic change irregularly as P_0 increases.

3. ELECTRON BUNCHING IN THE DRIFT SPACE FOR DIFFERENT OPERATING CONDITIONS OF TRAVELING-WAVE TUBES

Figure 3 shows the results of measurements of the power of the third harmonic signal along the electron beam emerging from the traveling-wave tube. The measurements were conducted at a beam current of 3.5 ma, a potential of the drift cylinder $U_g = U_c$, 500 v, and a magnetic field of 700 oe, for several operating conditions of the traveling-wave tube characterized by the signal power at the input P_0 (see Section 2). The abscissa in Figure 3 indicates the distance from the probe to the helix (in centimeters) while the ordinate shows the power of the third harmonic signal (in decimals in terms of 1 milliwatt). The dependence of the harmonic signal on the distance generally presents a very broken picture. Curves 3 and 4 are envelopes described through the maximum of the curve and characterize the main dependence of the

power of the harmonic on the bunching systems. When the signal at the input of the traveling-wave tube is less than 0.5 mw, the traveling-wave tube operates linearly and the power of the third harmonic in the beam is very small over the entire length of the beam and it is sufficient for clamping at the output of the receiver. Curve 1, Figure 3 corresponds to a power $P_0 = 0.5$ mw. The maximum values of the power of the harmonic occur at distances $l = 8-12$ cm. When the input signal is $P_0 = 2$ mw, the power of the harmonic is one order higher, and the maximum is shifted in the direction of greater distances.

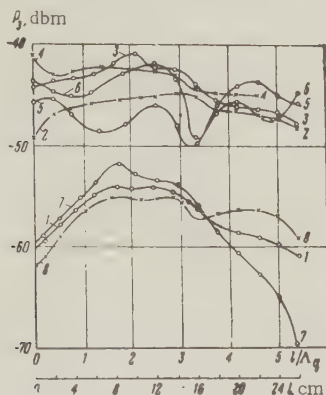


Figure 3. Dependence of the power of the third harmonics in electron beam at the output of the traveling-wave tube on the distance l along the beam for different power levels of the signal at the input of the tube.

$U_d = 500$ v: 1— $P_0 = 0.5$ mw; 2— $P_0 = 2$ mw; 3— $P_0 = 5.5$ mw; 4— $P_0 = 10$ mw; 5— $P_0 = 36$ mw; 6— $P_0 = 58$ mw; $U_d = 490$ v: 7— $P_0 = 0.5$ mw; $U_d = 550$ v: 8— $P_0 = 0.5$ mw.

Since the maximum of the harmonic must correspond approximately to the case of greatest bunching of electrons, the greatest bunching for $P_0 = 0.5$ mw occurs at distances 1 - 8 - 12 cm, while for $P_0 = 2$ mw ≈ 14 cm. With further increases of the input signal, the power of the harmonic increases and the maximum begins to shift toward smaller l . For $P_0 = 5.5$ mw, the maximum occurs at $l = 10$ cm. For $P_0 = 10$ mw, the curves of this power of the harmonic have several maxima (Curve 5 for $P_0 = 36$ mw and curve 6 for $P_0 = 58$ mw in Figure 3). This shape of the curves is the result of an amplification of the wave in the beam which is connected with electronic plasma oscillations.

Let us evaluate the distances l for Figure 3 in terms of the wavelength of the plasma oscillations of the space charge of electrons. On the basis of experimental, it is shown in Section 6 that the wavelength of the delayed plasma oscillations is given by the formula

$$\Lambda_q = 9.5 \cdot 10^{-3} U_d$$

for $P_0 = 36$ mw, $U_c = 500$ v, and a wavelength of the third harmonic $\lambda_3 = 9.5$ cm. For a drift potential $U_d = 500$ v, $\Lambda_q = 4.8$ cm and the frequency of the plasma operations is $f_q = 280$ Mc. A change in the potential U_d within the limits of 300 - 1300 v changes Λ_q within the limits of 2.9 - 12.3 cm. For the fundamental oscillations the wavelengths of the plasma oscillations is greater: for $U_d = 500$ v and $\lambda_0 = 15$ cm and $f_q = 90$ Mc. Thus the bunching of the electrons and the formation of the third harmonic, as described by the curves in Figure 3, occurs at distances somewhat greater than the wavelength of the plasma oscillations in the beam. At these distances, when the traveling-wave tube is in nonlinear operation ($P_0 > 2$ mw), the maximum value of the power of the harmonic in the plane of the exit and of the helix.

4. ELECTRON BUNCHING IN AN ACCELERATED OR DECELERATED BEAM

Figure 4 shows a graph giving the dependence of the signal power of the third harmonic on distance for potentials of the drift space not equal to the potential of the helix of the traveling-wave tube ($U_c = 500$ v). For small potentials U_d , the maximum of the harmonic power and, consequently, the formation of electron bunches occur at small distances from the helix (Curve 1: $P_0 = 0.5$ mw, $U_d = 100$ v; Curves 2, 7, 9, 11: $P_0 = 0.5 - 10$ mw, $U_d = 200$ v). As the potential of the drift space is increased, the region of existence of the harmonic is observed to shift in the direction of greater distances (Curves 6, 8 - 13 for $U_d = 900$ v, $P_0 = 0.5 - 36$ mw). This shift is seen more clearly in Curves 1 - 6 for $P_0 = 0.5$ mw, $U_d = 100 - 900$ v. The curves for the harmonic power as a function of distance (Figures 3 and 4) which were obtained for

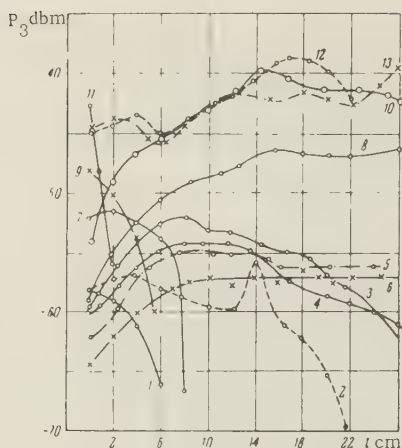


Figure 4. Dependence of the power of the third harmonic in electron beam at the output of the traveling-wave tube on the distance l along the beam for different potentials of the drift cylinder and different power levels of the signal at the output of the traveling-wave tube.

$P_0 = 0.5$ mw: 1— $U_d = 100$ v; 2— $U_d = 200$ v; 3— $U_d = 400$ v; 4— $U_d = 500$ v; 5— $U_d = 600$ v; 6— $U_d = 900$ v; $P_0 = 2$ mw: 7— $U_d = 200$ v; 8— $U_d = 900$ v; $P_0 = 5.5$ mw: 9— $U_d = 200$ v; 10— $U_d = 900$ v; $P_0 = 10$ mw: 11— $U_d = 200$ v; 12— $U_d = 900$ v; $P_0 = 36$ mw: 13— $U_d = 900$ v

operation of the traveling-wave tube with $P_0 \leq 10$ mw are characterized by one main maximum.

Each value of the power and the potential corresponds to a definite position in the maximum. The same conclusion can be drawn from oscillograms from the third harmonic signal as a function of the potential of the drift cylinder (Figures 5 and 6). For nearly linear operation of the traveling-wave tube as an amplifier, the oscillogram shows one major maximum standing out for many small peaks. This maximum shifts toward higher drift potentials as the probe is moved to greater distances. When the signals at the input of the traveling-wave tube are large ($P_0 > 10$ mw), the electron beam is strongly current- and velocity-modulated. On the oscillogram of Figure 6 which correspond to this type of beam, many maxima are observed. This is connected with the existence of plasma amplification of the oscillations in the beam (see Section 6).

The stability of the signal amplitude at the output of a frequency multiplier is connected with the brokenness of the curves on the oscillograms. The greatest stability, i.e., the least sensitivity to fluctuations of the voltage U_d , is obtained in operating conditions corresponding to least brokenness of the curve. This case corresponds, for instance, to the oscillograms of Figure 5 for P_0 equal 5.5 mw and $l = 0 - 6$ cm, and of Figure 6 for $P_0 = 10$ mw and $l = 0$. The least stability will occur under operating conditions characterized by oscillograms similar to the oscillograms in Figure 5 for $P_0 = 2$ mw and in Figure 6 for $P_0 = 36$ mw.

5. AMPLIFICATION OF WAVE FOR A RETARDED ELECTRON BEAM

It is shown in [12] that when the electron beam is decelerated, amplification of the ac waves and the velocity waves is observed. An experiment with a tube having a movable probe made it possible to trace the change of the amplification of the current wave of the third harmonic along the beam, which was retarded in the section containing the helix plus drift cylinder. The oscillograms of Figure 5 point to the existence of a rapid, almost discontinuous increase of the amplitude in the third harmonic as soon as the potential of the drift cylinder becomes smaller than the potential of the helix ($U_d < U_c = 500$ v). Amplification during deceleration is observed only under nearly linear conditions of operation of the traveling-wave tube. On the oscillograms, amplification appear most clearly at a power of the input signal $P_0 = 0.5$ mw and less clearly at $P_0 = 2$ mw (Figure 5). For $P_0 > 2$ mw, no large amplification is observed during deceleration. It can be seen from the oscillogram in Figure 5 that amplification during deceleration leads to an increase in the harmonic signal for potentials of the drift cylinder $300 \text{ v} < U_d < 500 \text{ v}$ at drift distances greater than the plasma wavelength in the beam. The greatest amplification of the power of the harmonic is 5 db (Figure 5) for $P_0 = 0.5$ mw and $l = 16$ cm. When the probe is moved to a distance $l > 16$ cm, the reverse is observed — the waves in the decelerated beam are attenuated. Characteristic oscillograms showing the appearance of a minimum are given in Figure 5 ($P_0 = 0.5$ mw, $l = 24$ cm; $P_0 = 2$ mw, $l = 16$ cm; $P_0 = 5.5$ mw, $l = 26$ cm).

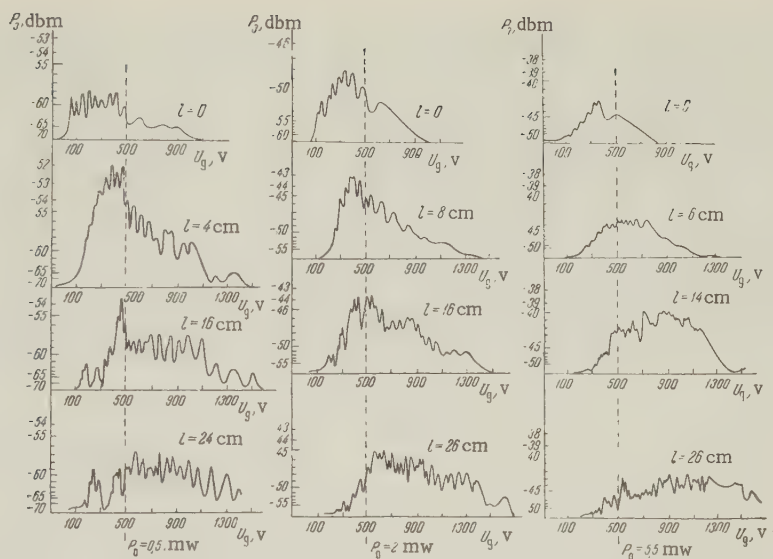


Fig. 5. Oscillograms of the third harmonic signal for signal powers of 0.5, 2, and 5.5 mw at the input of the traveling-wave tube.

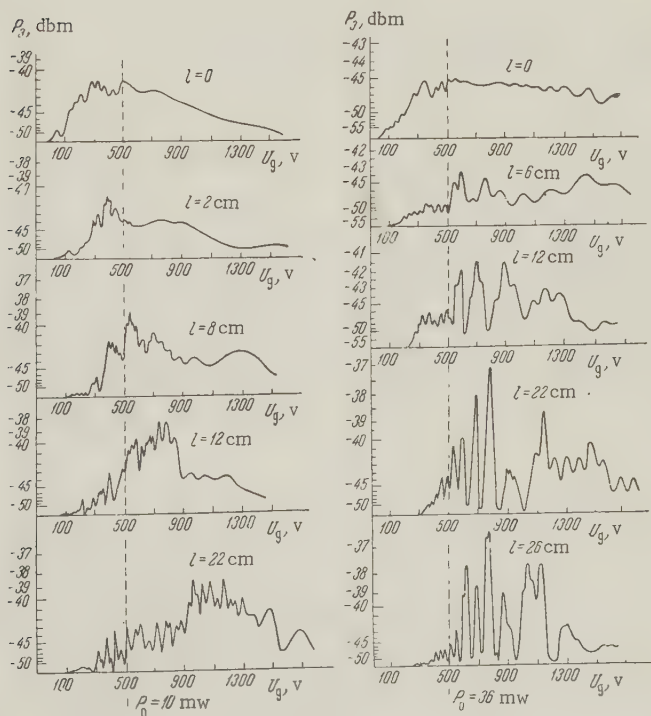


Fig. 6. Oscillograms of the third harmonic signal for signal powers of 10 and 36 mw at the input of the traveling-wave tube.

In an accelerated electron beam, attenuation of the amplitude of the harmonic occurs at small distances l and amplification occurs at L . The existence of attenuation explains the rapid decrease of the ordinate of the oscillogram curves in Figure 5 when the potential of the drift cylinder increases beyond 500 v, as well as the occurrence of a minimum of the harmonic signal in some region of the potential greater than 500 v. The minimum is clearly seen on the oscillogram for $P_0 = 0.5$ mw and $l = 16$ cm (Figure 5). Amplification of the harmonic signal for $U_d > 500$ v is observed on the oscillogram (Figure 5) for $P_0 = 2$ mw, $l = 16$ cm, and $P_0 = 5.5$ mw, $l = 26$ cm. The change of the power of the third harmonic along the beam for $P_0 = 0.5$ mw is shown in Figure 3 (Curve 1 for an unaccelerated beam $U_d = 500$ v, Curve 7 for a decelerated beam $U_d = 490$ v, Curve 8 for an accelerated beam $U_d = 550$ v). As compared with an unaccelerated beam, there exists amplification of the wave in the section $l = 0-17$ cm in a decelerated beam, while in an accelerated beam there is attenuation in the section. For $l > 17$ cm, the reverse is observed: attenuation of the waves in a deceleration beam, and amplification of the waves in an accelerated beam. The amplification of the harmonic signal in a retarding field can be used, for instance, in traveling-wave tube frequency multipliers, since the first stage of the multipliers operates nearly linearly [1, 2].

6. PLASMA AMPLIFICATION OF THE POWER OF THE THIRD HARMONIC ALONG THE BEAM FOR LARGE SIGNAL LEVELS AT THE INPUT OF THE TRAVELING-WAVE TUBE.

It was noted in Sections 3 and 4 that, at large signal levels at the input of the traveling-wave tube ($P_0 \geq 10$ mw), the power of the third harmonic is observed to increase with the distance in the beam. Let us examine the oscillograms of Figure 6 for $P_0 = 36$ mw. These oscillograms have peaks which depend quasi-periodically on the voltage U_d . These peaks appear, starting at a distance $l = 6$ cm, and the amplitude increases with increasing l . The periodicity of the recurrence of the peaks is disrupted somewhat by the existence of minute fluctuations of the oscillogram curves which are related to the pulsation of the diameter of the electron beam in a magnetic field. However the general nature of the periodic variation appears quite distinctly. The periodicity with respect to the voltage U_d is accompanied by a periodicity with respect to the position l , shown in Figure 7 for drift potentials $U_d = 700$ v and $U_d = 900$ v. The picture recalls a picture of standing waves. The length of a standing wave, equal to the distance between the peak maximum, increases with increasing potential of the drift space.

As the drift distance increases, the amplitude of the standing wave is seen to grow, i.e., the power of the third current harmonic increases. The greatest increase occurs for potentials U_d greater than the helix potential $U_c = 500$ v. The maximum of the harmonic signal appears at a potential $U_d = 800$ v and a drift distance $l = 23$ cm. The power of the harmonic at the maximum is approximately one order higher than the power of the harmonic in the plane of the end of the helix ($l = 0$). This phenomenon was studied for different values of the signal power at the input of the traveling-wave tube up to $P_0 = 150$ mw. The general nature of the periodic variation of the amplitude of the harmonic signal with distance and voltage was found to remain constant. The power of the harmonic was approximately the same as in the case $P_0 = 36$ mw. Let us note that for large signals ($P_0 = 10 - 150$ mw) the amplification in traveling-wave tube is close to unity, and the tube operates under the condition of the given field.

To explain the mechanism of the amplification of the harmonic signal, we conducted an analysis of the corresponding oscillograms and graphs to determine the wavelength of the oscillation producing a standing wave. For the case where the power of the signal at the input of the traveling-wave tube was equal to 36 mw, we obtained the relation $\Lambda = 9.5 \cdot 10^3 U_d$, where Λ is the required wavelength (in centimeters) and U_d is the potential of the drift cylinder (in volts). This relation is correct for small drift distances $l < 15$ cm. For $l > 15$ cm, Λ is observed to decrease with increasing l . The value of Λ depends on the power P_0 , and as this power increases Λ decreases. This is illustrated by the relation $\Lambda = 8.4 \cdot 10^{-3} U_d$, which is obtained for $P_0 = 58$ mw.

We evaluated the wavelengths of the electron plasma oscillations in the beam, taking into account the finiteness of the transverse dimensions Λ_q of the beam. For Λ_q , there exists the equation $\Lambda_q = \Lambda_p / p$ [14], where p is the reduction coefficient and Λ_p is the wavelength of the retarded plasma operation of an infinitely wide beam. $\Lambda_p = v_0 / f_p$ and $(2\pi f_p)^2 = \omega_p^2 = \eta j_0 / \epsilon_0 v_0$ where v_0 is the average velocity of the electrons, f_p is the

frequency of the plasma oscillation, η is the ratio of the charge of electrons to its mass, ϵ_0 is the permittivity of free space, and j_0 is the average current density. The value of j_0 can be determined only approximately, as 0.2 amp/cm² for a beam current $I_0 = 3.5$ ma. The values of the coefficients p for a ring-shaped beam in the drift cylinder are given in [13]. For the fundamental operation ($\lambda_0 = 25$ cm) at a drift potential of 500 v, we find $p = 0.15$. For the third harmonic ($\lambda_3 = 9.5$ cm) at $U_d = 500$ v, we have $p = 0.46$, while $p = 0.40$ at $U_d = 800$ v. The calculations made in a broad range of potentials U_d gave values of Λ_q approximately twice as large as the values of Λ . For instance, for the third harmonic at $U_d = 500$ v, we have $\Lambda = 4.8$ cm and $\Lambda_q = 11$ cm. However, it must be remembered that the values of Λ_q were found from the equations of a linear theory, which is correct for beams with small modulation of the current. Compared to such beams, the beam bunched in the drift cylinder of the output of the traveling-wave tube for large signal levels is quite nonlinear, and the frequency of the plasma oscillations in it is higher and the wavelength Λ_q smaller. Accordingly, it can be concluded that the standing waves of length Λ are the standing waves of the electronic plasma oscillations.

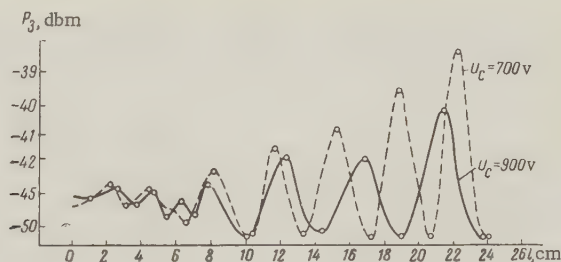


Fig. 7. The change of the power of the third harmonic with distance l along the beam for potentials of the drift cylinder of 700 and 900 v ($P_0 = 36$ mw).

In small signal operation, there are no standing waves of plasma oscillation (Figures 3-5). At the output of the traveling-wave tube, one slow space charge wave exists in the electron beam. In large signal operation, (of the given field), both space charge waves are excited, and a picture of standing waves with increasing amplitude is observed (Figures 6 and 7). The increase of amplitude is observed only under the conditions that the beam is sufficiently narrow, i.e., only if its transverse dimensions are smaller than the wavelengths of the retarded amplified oscillations (for $p < 0.5$ and $U_d > 500$ v). The phenomenon we are considering is apparently analogous to the increase of the amplitude of harmonics with distance in beams with initial oscillation of velocity (described in articles [14, 15]). Indeed, in the case of velocity modulation, the increase of the amplitude of the harmonics occurs only in narrow electron beams and is manifested as an increase of the amplitude of the standing wave of the electronic plasma oscillations. In the above-mentioned articles, the phenomenon is treated as parametric amplification with low-frequency pumping at the frequency of the fundamental component of current.

CONCLUSIONS

1. The method of extracting the energy of harmonics which was used in the experimental tube to study the phenomenon of electron bunching and which is based on the radiation of electron bunches near inhomogeneities is simple and sufficiently sensitive.

2. The maximum values of the power of the third harmonic are obtained in distances greater than the wavelength of the retarded plasma oscillations in the beam. During small signal operation at the traveling-wave tube input, the dependence of the harmonic power on the distance and on the potential of the drift space is characterized by one major maximum and does not contain standing waves due to modulation by plasma oscillations.

3. The maximum values of the power of the third harmonic of the bunched beam in the drift space are smaller than the maximum value of the harmonic power in the traveling-wave tube. However the difference does not exceed 4 db.

4. By changing the potential of the drift space, it is possible to shift the region of existence of the harmonics to different distances from the helix of the traveling-wave tube.

5. For a large signal level at the input of the traveling-wave tube, plasma amplification of the waves appear in the beam. This leads to an additional increase of the power of the harmonic signal by approximately one order.

6. At potentials of the drift space smaller than the potential of the helix of the traveling-wave tube, the deceleration of electrons in the section containing the helix plus drift cylinder makes it possible to obtain additional amplification of the harmonic signal in the beam.

7. The oscillograms showing the variation of the harmonic signal with the drift potential make it possible to estimate the stability of the operation of the multiplier with respect to fluctuations of the potential under different operating conditions.

REFERENCES

1. V.I. Kanavets, Two-beam frequency multiplier using space-separated helices, Report presented at the All-Union Conference on Radioelectronics, Saratov, 1957.
2. D.J. Bates, E.L. Ginzton, Traveling-wave frequency multiplier, Proc. I.R.E., 1957, 45, 7, 938.
3. R.W. De Grasse, D.A. Dunn, R.W. Grow, G. Wide, Microwave frequency mixing and frequency division with beam-type tubes, IRE Wescon Convention Record, 1957, 1, 3, 163.
4. A.J. Ashkin, Parametric amplification of space charge wave, J. Appl. Phys., 1958, 29, 12, 1646.
5. V.B. Braginskiy and Ye.R. Mustel', Experimental study of the radiation of electron beams near inhomogeneities, Izv. Vuzov MVO SSSR (Radiofizika), 1958, 1, 3, 124.
6. A. Nordsieck, Theory of the large signal behavior of traveling-wave amplifiers, Proc. I.R.E., 1953, 41, 5, 630.
7. L.A. Vaynshteyn, Nonlinear theory of traveling-wave tubes, Radiotekhnika i Elektronika, 1957, 2, 7, 883; 1957, 2, 8, 1027.
8. I.E. Rowe, A large-signal analysis of the traveling-wave amplifier: theory and general results, IRE Trans. Electron Devices, 1956, ED-3, 1, 39.
9. Klystrons, Izd. Sovetskoye Radio, 1952.
10. S.E. Webber, Large signal analysis of the multicavity klystron, IRE Trans. Electron Devices, 1958, ED-5, 4, 98.
11. W.W. Rigrod, Noise spectrum of electron beam in longitudinal magnetic field, Bell System Techn., J., 1957, 37, 4, 831.
12. P.K. Tien, L.M. Field, Space-charge waves in an accelerated electron stream for amplification microwave signals, Proc. I.R.E., 1952, 40, 6, 688.
13. G.M. Branch, T.G. Mihran, Plasma frequency reduction factors in electron beams, IRE Trans. Electron Devices, 1955, ED-2, 2, 3.
14. T.G. Mihran, Harmonic current growth in velocity-modulated electron beams, J. Appl. Phys., 1959, 30, 9, 1346.
15. F. Paschke, Generation of second harmonic in a velocity-modulated electron beam of finite diameter, RCA Rev., 1958, 19, 4, 617.

Department of Physics
Moscow State University
im. M. V. Lomonosov,
Chair of Radio Engineering

Received May 21, 1960

FOCUSING THE ELECTRON BEAM IN A TRAVELING-WAVE TUBE BY A PERIODIC ELECTROSTATIC FIELD

A. L. Igritskiy

A method is proposed for computing the trajectories of electrons in solid or hollow electron beams in a traveling-wave tube with a periodic electrostatic field produced by bifilar spirals.

A differential equation for the trajectory of an electron is derived and it is integrated for the more difficult case — that of a hollow beam. With certain changes, the equations obtained are applicable to solid beams.

The condition for optimal focusing is derived. When this condition is satisfied, the boundary trajectories of the hollow beam are approximately parallel to the axis of the tube, or (the case of a solid beam) the undulation of electron trajectories at a given radius is small.

It is shown that a periodic electrostatic field can be a "transformer" of the cross section of the hollow beam. It is demonstrated that periodic electrostatic focusing of a solid beam is also possible in the case where an ordinary electron gun with a beam of electrons with uniform velocity is used.

1. DERIVATION OF THE DIFFERENTIAL EQUATION FOR THE TRAJECTORY OF AN ELECTRON IN A HOLLOW BEAM AND ITS INTEGRATION

To focus a hollow beam, a biperiodic focusing device [1] is used, the circuit of which is given in Figure 1. This device consists of two bifilar spirals, between which the ring-shaped beam passes. For simplicity, we shall consider that the two bifilar spirals have the same pitch L and the same potential difference $2V_f$ between strips. The average potential of both spirals is equal to V_0 . The potential distribution in the space between the spirals is determined for this case in [2], and is represented by an infinite series. If we retain only the first term of this series, the other terms being small, we can write for the potential distribution

$$V(r; \varphi; z) = V_0 + V(r) \cos\left(\frac{2\pi}{L}z - \varphi\right), \quad (1)$$

where

$$\begin{aligned} V(r) = & \\ = \frac{4V_f \sin \pi\sigma}{\pi^2\sigma} \frac{\left[K_1\left(\frac{2\pi}{L}r_2\right) - K_1\left(\frac{2\pi}{L}r_1\right) \right] I_1\left(\frac{2\pi}{L}r\right) - \left[I_1\left(\frac{2\pi}{L}r_2\right) - I_1\left(\frac{2\pi}{L}r_1\right) \right] K_1\left(\frac{2\pi}{L}r\right)}{I_1\left(\frac{2\pi}{L}r_1\right) K_1\left(\frac{2\pi}{L}r_2\right) - I_1\left(\frac{2\pi}{L}r_2\right) K_1\left(\frac{2\pi}{L}r_1\right)} = & \\ = 2\xi(r) V_f. & \end{aligned} \quad (2)$$

In Eqs. (1) and (2), σ is the ratio of the distance d between adjacent turns of the spiral to the period L of the electrostatic field, which is equal to the pitch of the bifilar spiral; I_1 and K_1 are modified first-order Bessel functions of the first and second kind; r_1 is the outer radius of the inner spiral; r_2 is the inner radius of the outer bifilar spiral; and φ is the angle of rotation about the z -axis, which is the axis of symmetry of the traveling-wave tube. In the given case, the electric field depends on the angle φ and is not axially symmetric. Therefore, it is necessary to consider the motions of the electrons in the electrostatic field along the three cylindrical coordinates, r , φ and z .

These equations have the form [3]

$$\ddot{r} - r\dot{\varphi}^2 - \eta \frac{\partial V}{\partial r} = -\eta E_r, \quad (3)$$

$$r\ddot{\varphi} + 2\dot{r}\dot{\varphi} = \frac{1}{r} \eta \frac{\partial V}{\partial \varphi}, \quad (4)$$

$$\ddot{z} = \eta \frac{\partial V}{\partial z}. \quad (5)$$

Here the dots indicate derivatives with respect to time; V is the electrostatic potential determined from Eq. (1); η is the ratio of the electron charge to its mass; E_r is the radial component of the intensity of the electric field produced by the space charge of the hollow beam. This radial component can be determined from the equation [1]

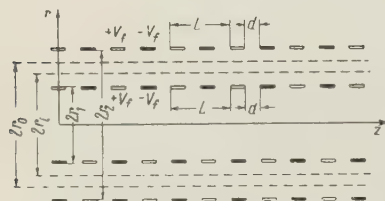


Fig. 1. Circuit of a biperiodic focusing device

$$E_r = -\frac{dV}{dr} = -\frac{j}{2e_0 v_z} (r^2 - r_e^2), \quad (6)$$

where j is the density of the current in the beam; v_z is the axial velocity of the electron; ϵ_0 is the permittivity of free space; and r_e is the equilibrium radius at which the field of the space charge is equal to zero.

We shall consider that the strips of the inner and outer bifilar spirals which are located opposite one another are at the same potential. Then the equilibrium radius can be found from the equation [1]

$$r_e^2 = \frac{1}{\ln \frac{r_0}{r_1}} \left[\left(r_0^2 \ln \frac{r_2}{r_0} + r_i^2 \ln \frac{r_i}{r_1} \right) + \frac{1}{2} (r_0^2 - r_i^2) \right], \quad (7)$$

where r_0 and r_1 are the outer and inner radii of the hollow beam respectively.

Let the current be distributed uniformly across the beam and suppose that the electron trajectories do not intersect in the process of motion. Then, from eq. (6), in terms of the absolute value,

$$|j(r^2 - r_e^2)| = \frac{I_r}{\pi} = \text{const}, \quad (8)$$

where I_r is the part of the current passing in the space between the cylinder of radius r_e and the cylindrical surface obtained by rotating about the z -axis the trajectory of an electron which passes through a point with coordinates $(r; z)$. This current will be constant, since no electrons pass in or out of the lateral surface of the above-defined space. If the value of E_r is substituted from (6) into (3), a nonlinear differential equation will be obtained. To make this equation linear, we shall consider small changes of the radius of the beam about some value $r_v = \text{const}$, equal to the radial coordinates of the electron in entering the periodic device:

$$r = r_v(1 + \delta), \quad (9)$$

where $\delta \ll 1$. To simplify the problem, we shall also assume that the angular velocity of the electron φ resulting from the asymmetry of the field is a small quantity close to zero. Under this assumption we can neglect the higher order term $r\varphi^2$ in eq. (3). Let us expand the potential $V(r)$ in eq. (2) into a Taylor series of powers of $r_v\delta$ in the neighborhood of $r = r_v$:

$$V(r) = V(r_v) + V'(r_v)r_v\delta + \frac{1}{2}V''(r_v)r_v^2\delta^2 + \dots \quad (10)$$

Here the primes indicate derivatives with respect to r . Using (6), (8), (9), (1), and (10), in combination with the above remarks, Eq. (3) can be written in the form

$$\dot{\delta} + \left[\pm \frac{\eta I_r}{2\pi e_0 v_z r_V^2} - \eta V''(r_V) \cos\left(\frac{2\pi}{L} z - \varphi\right) \right] \delta = \pm \frac{\eta I_r}{2\pi e_0 v_z r_V^2} + \frac{\eta V'(r_V)}{r_V} \cos\left(\frac{2\pi}{L} z - \varphi\right). \quad (11)$$

In the terms of Eq. (11) which have a double sign, the plus sign corresponds to the trajectory of an electron which enters the periodic field with a radial coordinate $r_V > r_e$, or the minus sign corresponds to an electron which enters the field with the coordinate $r_V < r_e$. Considering the cross section of the electron beam at the entry to the periodic electrostatic field, let us express the current I_r in terms of the total current I_f of the hollow beam, and let us define

$$B_V^2 = \frac{V^2 I_f}{\pi e_0 \eta^{1/2} V_0^2 r_V^2} \left| \frac{r_V^2 - r_e^2}{r_0^2 - r_i^2} \right|. \quad (12)$$

To transform to the differential equation of the trajectory, let us replace the derivatives with respect to time in (11) by the derivative with respect to the z -coordinate. In the given case the electron velocity changes along the z -axis as a result of the periodic electrostatic field; it follows that

$$\ddot{\delta} = v_z^2 \frac{d^2 \delta}{dz^2} + v_z \frac{dv_z}{dz} \frac{d\delta}{dz}. \quad (13)$$

We will then obtain an equation which will contain the first derivative $d\delta/dz$. To eliminate the term with the first derivative, let us use a function y defined by

$$\delta = v_z^{-1/2} y. \quad (14)$$

As a result of the transformations, we have

$$\begin{aligned} \frac{d^2 y}{dz^2} + \left[\frac{1}{4v_z^2} \left(\frac{dv_z}{dz} \right)^2 - \frac{1}{2v_z} \frac{d^2 v_z}{dz^2} \pm \frac{\eta B_V^2}{8V_0} \left(\frac{v_0}{v_z} \right)^3 - \frac{\eta V''(r_V)}{r_z^2} \cos\left(\frac{2\pi}{L} z - \varphi\right) \right] y = \\ = \pm \frac{\eta B_V^2}{8V_0} v_0^3 v_z^{-5/2} + \frac{\eta V'(r_V)}{r_V} v_z^{-3/2} \cos\left(\frac{2\pi}{L} z - \varphi\right). \end{aligned} \quad (15)$$

The electron velocity v_z can be expressed in terms of the potential, and if we take into account (1) and (10), it will be equal to

$$v_z = \sqrt{2\eta V} = v_0 \left[1 + \frac{V(r_V)}{2V_0} \cos\left(\frac{2\pi}{L} z - \varphi\right) \right]. \quad (16)$$

In the derivation of Eq. (16), it is assumed that $V(r_V)/V_0 \ll 1$. The velocity v_0 can be determined from (16) by setting $V = V_0$. We shall regard the φ in the cosine term as constant, taking it to be equal to the value of φ_V at the entry to the periodic field: $\varphi = \varphi_V = \text{const}$. If we change the independent variable by the substitution

$$t = \frac{\pi}{L} z - \frac{\varphi_V}{2} \quad (17)$$

and substitute (16) into (15), after a series of transformations we will finally obtain the following differential equation:

$$\frac{d^2 y}{dt^2} + (a - 2q \cos 2t) y = b + 2p \cos 2t. \quad (18)$$

In equation (18), the following notation is used:

$$a = \pm \frac{\eta B_V^2}{8V_0} \left(\frac{L}{\pi} \right)^2, \quad (19)$$

$$2q = \left(\frac{L}{\pi}\right)^2 \frac{V''(r_v)}{2V_0} - \frac{V(r_v)}{V_0} \pm \frac{3}{16} \frac{\eta B_v^2 V(r_v)}{V_0} \left(\frac{L}{\pi}\right)^2, \quad (20)$$

$$b = \pm \left(\frac{L}{\pi} \right)^2 \frac{\eta B_v^2}{8V_0} v_0^{1/2}, \quad (21)$$

$$2p = \left(\frac{L}{\pi}\right)^2 v_0^{1/2} \left[\frac{V'(r_V)}{2V_0 r_V} \mp \frac{5}{32} \frac{\eta B_V^2}{V_0} \frac{V(r_V)}{V_0} \right]. \quad (22)$$

Equation (18) is an inhomogeneous Mathieu differential equation. The coefficients of the equation, defined by formulas (19) - (22) are constant for the given electron trajectory with an entry radius $r = r_v$. However if the entire hollow beam is considered, r_v will change in it from r_0 to r_1 , and accordingly the coefficient of the equation will change. As known from [4], a Mathieu equation can have stable and unstable solutions. The nature of the solution is determined from a stability diagram, given in Figure 2, for the coefficients a and q . It is seen from Eqs. (12) and (19) as the entry radius r_v changes from r_0 to r_1 the coefficient a will change from a positive value a_n to some negative value a_1 , and that at $r_v = r_e$ it will be equal to zero. The coefficient $2q$ also varies as r_v changes. This variation of the coefficient a makes it possible to conclude that a stable hollow beam will be obtained if all the electron trajectories correspond mathematically to the zero stability zone of the Mathieu equation $n = 0$, found by the curves a_0 and b_1 .

The general solution of an inhomogeneous Mathieu equation is given in article [5]. For simplicity, we shall consider the trajectory of an electron which enters a periodic field at a maximum with a radial coordinate $r = r_v$, moving in the plane $\varphi = 0$ parallel to the axis of the tube. The initial conditions can then be written in the form (for $z_e = t_0 = 0$)

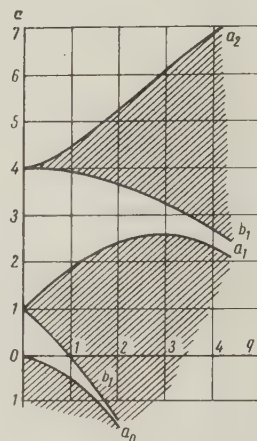


Fig. 2. Stability diagram for Mathieu functions

$$\varphi_v = 0, \delta_v = y_v = 0, y'_v = r'_v = \delta'_v = 0. \quad (23)$$

According to [5], the solution of Equation (18) for these initial conditions with $n = 0$ will be as follows:

$$y(t) = y_{c0} \cos \beta t + y_{c-2} \cos [(\beta - 2)t] + y_{c+2} \cos [(\beta + 2)t] + y_n + y_2 \cos 2t + y_4 \cos 4t. \quad (24)$$

In Eq. (24), the following notation is used

$$y_{c0} = \frac{1}{c^2} [c_0^2 (A_0 - B_0 - D_0) + c_0 c_{-2} (A_{-1} - B_{-1} - D_{-1}) + c_0 c_{+2} (A_1 - B_1 - D_1)], \quad (25)$$

$$y_{c-2} = \frac{1}{c^2} c_0 c_{-2} (A_0 - B_0 - D_0), \quad (26)$$

$$y_{c+2} = \frac{1}{c^2} c_0 c_{+2} (A_0 - B_0 - D_0), \quad (27)$$

$$y_{\text{II}} = \frac{1}{c^2} [-c_0^2 A_0 + c_0 c_{-2} (B_0 + D_{-1}) + c_0 c_{+2} (B_1 + D_0)], \quad (28)$$

$$y_2 = \frac{1}{c^2} [c_0^2 (B_0 + D_0) - c_0 c_{-2} (A_0 + A_{-1}) - c_0 c_{+2} (A_0 + A_1)], \quad (29)$$

$$y_4 = \frac{1}{c^2} [c_0 c_{-2} (D_0 + B_{-1}) + c_0 c_{+2} (B_0 + D_1)], \quad (30)$$

$$c^2 = c_0^2\beta + 2[c_0c_{-2}(\beta - 1) + c_0c_{+2}(\beta + 1)], \quad (31)$$

$$\beta \approx \left\{ a - \frac{a-1}{12(a-1)^2 - q^2} q^2 - \frac{5a+7}{32(a-1)^3(a-4)} q^4 \right\}^{1/2}. \quad (32)$$

The coefficients c_0 , c_{-2} , and c_{+2} are found from the recursion formula

$$[a - (2r + \beta)^2] c_{2r} - q(c_{2r-2} + c_{2r+2}) = 0. \quad (33)$$

The coefficients, A, B and D which appear in Eqs. (25) - (30), depend on the inhomogeneous part of the Mathieu equation. In the given case, according to [5],

$$\begin{aligned} A_0 &= -\frac{b}{\beta}, \quad B_0 = \frac{p}{\beta-2}, \quad D_0 = \frac{p}{\beta+2}, \\ A_{-1} &= -\frac{b}{\beta-2}, \quad B_{-1} = \frac{p}{\beta-4}, \quad D_{-1} = \frac{p}{\beta}, \\ A_1 &= -\frac{b}{\beta+2}, \quad B_1 = \frac{p}{\beta}, \quad D_1 = \frac{p}{\beta+4}. \end{aligned} \quad (34)$$

If we have an equation for y , the undulation of the beam can be determined from (14). For this it is sufficient to set $v_z = v_0$.

2. EXAMPLE

Let us compute the configuration of a hollow beam in a traveling-wave tube by periodic focusing. The electrostatic field is produced by two bifilar spirals, having the following dimension (see Figure 1): $r_1 = 3$ mm, $r_2 = 5$ mm, $L_1 = L_2 = L = 2.8$ mm, $d_1 = d_2 = d = 0.8$ mm. The potential difference between terms is identical for the two spirals and is equal to $2V_f = 550$ v. In the electron gun, the electrons are accelerated by a constant voltage $v = 1800$ v. The beam current is $I_b = 4 \cdot 10^{-3}$ amp. At the entrance to the periodic electrostatic field, the beam has the dimensions $r_0 = 4.2$ mm, $r_i = 3.8$ mm. The outer boundary of the beam is determined by the trajectory of an electron having a radial coordinate $r_v = r_0$ at the entrance. Computing the coefficients appearing in (18), we obtain

$$\frac{d^2 y}{dt^2} + (2 \cdot 10^{-5} - 3.58 \cdot 10^{-2} \cos 2t) y = 0.1 + 19.4 \cos 2t. \quad (35)$$

The computed values of a and q in Figure 2 correspond to the point in the zone of stability $n = 0$. By determining the coefficients from (25) - (34), and by substituting their values into (24), we obtain $y(t)$. In view of Eqs. (14) and (9), the outer boundary of the beam will be described by the equation

$$\begin{aligned} r_{o.b.} &= 4.2 \left[1 - 13.5 \cdot 10^{-3} \cos \left(\frac{2\pi}{418} z \right) + 0.498 \cdot 10^{-3} \cos \left(\frac{2\pi}{2.82} z \right) + \right. \\ &\quad \left. + 0.486 \cdot 10^{-3} \cos \left(\frac{2\pi}{2.78} z \right) + 14.4 \cdot 10^{-3} - 1.884 \cdot 10^{-3} \cos \left(\frac{2\pi}{2.8} z \right) \right]. \end{aligned} \quad (36)$$

In Eq. (36), r and z are expressed in millimeters.

The inner boundary of the beam is determined by the trajectory of an electron having a radial coordinate $r_v = r_i$ at the entrance. This trajectory is computed in a similar fashion to the preceding one, but the lower sign is taken in the terms with double signs in Eqs. (19) - (22). In the given case, eq. (18) will have the form

$$\frac{d^2 y}{dt^2} + (-1.92 \cdot 10^{-5} - 4.17 \cdot 10^{-2} \cos 2t) y = -9.65 \cdot 10^{-2} - 16.1 \cos 2t. \quad (37)$$

The computed values of a_i and q_i in Figure 2 correspond to a point located below the abscissa. Because of the small magnitude of the quantities a_i and q_i , it is difficult to establish from the graph whether the solution is stable. It follows from the theory of Mathieu functions [4] that the boundary of stability in the given case is determined by the eigenvalue of the function $ce_0(z; q)$ and is represented by the equation

$$a_0 \simeq -\frac{1}{2} q^2 + \frac{7}{128} q^4. \quad (38)$$

If we substitute the value $q_1 = 2.085 \cdot 10^{-2}$, we will obtain $a_0 = 2.175 \cdot 10^{-4}$. We see that in terms of absolute values $|a_0| > |a_1|$. Consequently, the solution of the Mathieu equation (37) will be stable. However, since the operating point on the stability diagram is close to the curve a_0 , the reserve stability of the inner boundary of the trajectory is small. Indeed, the disruption of stability, as in increasing the current, seems to occur on the inner boundary of the beam.

Equation (38) makes it possible to explain the experimental results of Chang [1, 6], namely the transconductance of $K = I/V_0^{3/2}$ at low accelerating voltages. According to Equations (12) and (19), as the voltage V_0 decreases, the absolute magnitude of the coefficient a increases proportionally to $1/V_0^{3/2}$. According to Equation (20) and (38), the eigenvalue of the function $ce_0(z; q)$, which determines the stability, increases more rapidly, proportionally to $1/V_0^2$, as the voltage decreases. Thus, as the voltage decreases the reserve of stability grows and the transconductance K increases.

The equation for the inner boundary of the beam has the form

$$r_{ib} = 3.8 \left[1 + 11.45 \cdot 10^{-3} \cos\left(\frac{2\pi}{397} z\right) - 0.506 \cdot 10^{-3} \cos\left(\frac{2\pi}{2.82} z\right) - 0.49 \cdot 10^{-3} \cos\left(\frac{2\pi}{2.78} z\right) - 12.17 \cdot 10^{-3} + 1.712 \cdot 10^{-3} \cos\left(\frac{2\pi}{2.8} z\right) \right]. \quad (39)$$

From Eqs. (36) and (39), the configuration of the hollow beam in the biperiodic focusing device is constructed in Figure 3. This figure also shows the trajectory of an electron having an entry radius $r_v = 4.1$ mm. The electron trajectories do not intersect and the motion of electrons in the beam is approximately laminar. It is also seen from Figure 3 that considerable deflection of electrons in the radial direction occurs in the process of motion in the periodic electrostatic field. In this connection the question arises of whether it is possible to select the parameters of the focusing device in a way suitable for producing an electron beam parallel to the axis of the tube.

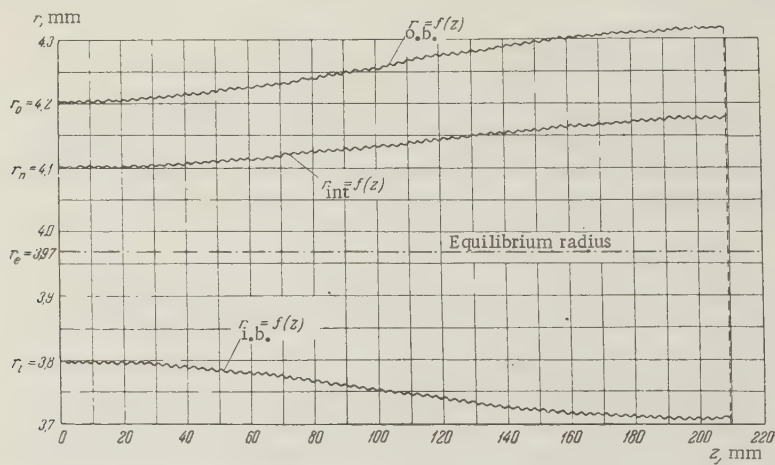


Fig. 3. Configuration of the hollow beam in the biperiodic focusing device ($r_{o.b.} = f(z)$ is the outer boundary of the beam, $r_{i.b.} = f(z)$ is the inner boundary of the beam, $r_{int} = f(z)$ is the intermediate trajectory with $r_v = 4.1$ mm)

3. METHOD OF PRODUCING A HOLLOW BEAM PARALLEL TO THE TUBE AXIS

An analysis shows that the most important contribution to the radial deflection of an electron is made by terms containing y_{c0} and y_p . A characteristic property of the Mathieu equation for the case of electrostatic focusing of beams is the small value of the coefficients a and q (a and $q \ll 1$). This makes it possible to derive simple approximate equations for y_{c0} and y_p .

In view of the smallness of the quantities a and q , Equation (32) can be written with sufficient accuracy in the form

$$\beta \simeq \left(a + \frac{q^2}{2}\right)^{1/2}. \quad (40)$$

Because of the smallness of the coefficients a and q , $\beta \ll 1$. In computing the coefficients c_{+2} and c_{-2} from (33), the coefficients c_{-4} and c_{+4} can be neglected. As a result, we obtain the following equation for c_{-2} and c_{+2} :

$$c_{-2} = c_{+2} \simeq -\frac{q}{4} c_0. \quad (41)$$

Because of the smallness of β , Equations (34) for the coefficients A , B , and D can be rewritten in the form

$$\begin{aligned} A_0 &= -\frac{b}{\beta}, \quad B_0 = -\frac{p}{2}, \quad D_0 = \frac{p}{2}, \\ A_{-1} &= \frac{b}{2}, \quad B_{-1} = -\frac{p}{4}, \quad D_{-1} = \frac{p}{\beta}, \\ A_1 &= -\frac{b}{2}, \quad B_1 = \frac{p}{\beta}, \quad D_1 = \frac{p}{4}. \end{aligned} \quad (42)$$

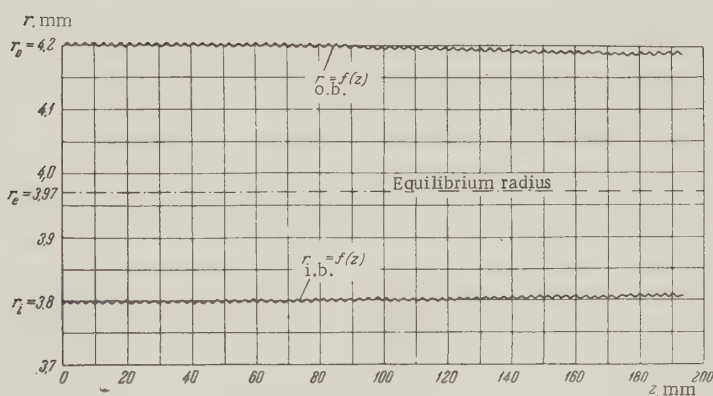


Fig. 4. Configuration of the hollow beam in the biperiodic focusing device at the optimal voltage between the spiral turn ($r_{o.b.} = f(z)$ is the outer boundary of the beam, and $r_{i.b.} = f(z)$ is the inner boundary of the beam.

Substituting these values into Eqs. (25), (28) and (31), we find

$$y_n = -y_{c0} = \frac{1}{\beta^2} \left[b - \frac{qp}{2} \right], \quad (43)$$

and Equation (24) can be written in the approximate form

$$y(t) = \frac{1}{\beta^2} \left(b - \frac{qp}{2} \right) (1 - \cos \beta t). \quad (44)$$

It is desirable that the radial deflection of the boundary trajectory of the beam be as small as possible. By setting

$$b = \frac{qp}{2}, \quad (45)$$

we obtain an almost parallel electron beam.

From Eq. (45), it is possible to determine the optimal potential difference between the turns of the bifilar spiral which will provide a trajectory in the outermost electron parallel to the axis. For this we substitute the values of the quantities defined by Eqs. (20) - (22) into (45). The computation shows that in these equations for q and p, the last terms are only small corrections, 2 to 3 orders of magnitude smaller than the preceding terms. Therefore they can be neglected. Thus, we obtain

$$(2V_f)_{\text{opt}} = \sqrt{\frac{2r_v V_0 \eta (\pm B_v^2)}{\zeta'(r_v)} \left[\frac{L^2 \zeta''(r_v)}{\pi^2} - \zeta(r_v) \right]^{-1}} \tag{46}$$

The coefficients ζ , ζ' and ζ'' are determined from (2). Figure 4 shows the trajectories of the boundary electrons of the beam computed for the optimal voltage between the turns of the spiral (equal in our example to $2V_f = 592$ v). We see that the boundary trajectories of the beam are parallel to the axis of the tube.

By considering Figures 3 and 4, it is easily seen that a periodic electrostatic field whose length along the z-axis is L/β can serve as a special transformer of the cross section of a hollow beam. Indeed, the cross section of the beam at the outlet of the periodic field can be made larger, smaller, or equal to the cross section at the entrance. The electrons which enter the periodic electrostatic field traveling parallel to the axis of the system at the exit will also lie approximately parallel to the axis. For the cross section of the beam at the exit to be equal to the cross section at the entrance, it is obviously necessary that the optimal voltage be applied between the strips of the spiral (see Figure 4). If the voltage between the strips of the spiral is somewhat smaller than the optimal voltage, the cross section of the beam at the exit will be greater than at the entrance (see Figure 3), and vice versa.

4. FOCUSING A SOLID BEAM WITH A BIFILAR SPIRAL

Figure 5 shows a bifilar ribbon spiral for periodic electrostatic focusing of a solid electron beam [7]. The potential distribution inside the spiral can be computed from the equation [2]

$$V(r; \varphi; z) = V_0 + V(r) \cos\left(\frac{2\pi}{L} z - \varphi\right), \tag{47}$$

where

$$V(r) = 4V_f \frac{\sin \pi z}{\pi^2 z} \frac{I_1\left(\frac{2\pi}{L} r\right)}{I_1\left(\frac{2\pi}{L} r_a\right)} = 2\zeta(r) V_f. \tag{48}$$

Here r_a is the inner radius of the bifilar spiral, while the other notations are the same as above. The coefficient B_v^2 for a solid beam is determined from the equation [7]

$$B_v^2 = \frac{V \sqrt{2} I}{\pi \epsilon_0 \eta^2 \omega^2 r_a^2}. \tag{49}$$

Differential equation (18) will be correct for the trajectory of an electron. In Eqs. (19) - (22) for the coefficient of this equation, only the upper sign should be taken in terms with double signs. The values of the potential, its derivatives and B_v^2 which appear in the equations should be substituted from Eqs. (47) - (49). Because of the small magnitude of the period L in the case of electrostatic focusing, the coefficients a and q are small, and therefore these values of a and q correspond to the zero stability zone $n = 0$ on the stability diagram (Fig. 2). We shall consider trajectories for which the initial condition (23) are satisfied. The solution of Eq. (18) for these initial conditions with $n = 0$ will be determined, as in the case of a hollow beam, by Equations (24) - (34).

These equations were used to compute the trajectory of the outermost electron in a

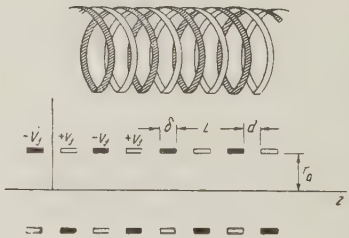


Fig. 5. Bifilar ribbon spiral (the ordinate indicates r).

traveling-wave tube with a periodic electrostatic field produced by a bifilar spiral having the following dimensions (see Figure 5): $r_a = 2$ mm, $L = 3.6$ mm, and $d = 1.04$ mm. The potential difference between the strips of the spiral is $2V_f = 170$ v. In the electron gun, the electrons are accelerated by a constant voltage $V_0 = 500$ V. The current of the electron beam is $I = 10^{-3}$ amp. At the entrance to the periodic electrostatic field, the beam has a radius $r_0 = 1.6$ mm. The computed trajectory of the outermost electron is shown in Figure 6 (curve a). We see that in the process of motion in a periodic field, there is a considerable deflection of electrons in the radial direction.

The undulation of the trajectory will be minimum if, as in the case of hollow beams, equation (46) is satisfied. Figure 6 also shows the trajectory of the outermost electron for the optimal voltage between the strips of the spiral $2V_f = 202$ v, computed from (46), (Curve b). It is easily seen that the radial deflection of this trajectory is small.

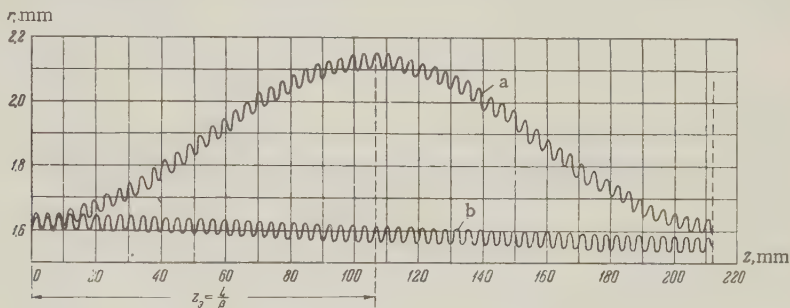


Fig. 6. Trajectory of the outermost electron $r = f(z)$ in the bifilar spiral: a — for a nonoptimal voltage between the strips of the spiral, $2V_f = 170$ v; b — for the optimal voltage between the strips of the spiral $2V_f = 202$ v.

The method of computing trajectories considered above is correct not only for peripheral electrons with an entrance radius $r_v = r_0$, but also for electrons entering the periodic field at an arbitrary distance r_v from the axis. For this it is necessary to replace the total current I in the equation by the current $I(r_v)$ flowing inside a cylinder of radius r_v . Since the coefficients of the Mathieu equation $2q$ and $2p$ depend strongly on the entrance radius r_v , it can be concluded that the motion of an electron with an entrance radius r_{v1} will differ from the motion of a peripheral electron. Thus, intersection of the electron trajectories becomes possible, and an electron emerging from the edge of the cathode will not always remain on the edge.

From Eq. (45), it is possible to define the function $V_0 = f(r)$ to provide small undulation of electron trajectories in the beam. This function is constructed in Figure 7. The development of an electron gun with a radial distribution of the accelerating potential toward Figure 7 is a complicated problem. Therefore, it is more advisable to use hollow beams having a small thickness in the radial direction.

Still another method can be proposed for reducing the radial deflection. It can be seen from Figure 6 that the extreme value of the radial coordinate of the electron occurs at

$$z_e = \frac{L}{\beta} = \frac{L}{\left[a + \frac{q^2}{2} \right]^{1/2}} = \frac{\pi}{\left\{ \frac{\eta B_v^2}{8V_0} + \frac{1}{2} \left[\frac{V''(r_v)}{4V_0} - \left(\frac{\pi}{L} \right)^2 \frac{V(r_v)}{2V_0} \right] \right\}^{1/2}}. \quad (50)$$

The maximum radial deflection of the electrons in the spiral of the tube can obviously be reduced if z_e is considerably greater than the length of the spiral: $z_e \gg 1_{sp}$. From Eq. (50), it can be seen that it is possible to increase z_e by using a greater accelerating voltage V_0 .

Figure 8 shows electron trajectories for different entrance radii, obtained as a result of computation with Eqs. (44), (14), and (9) for a beam with a bifilar spiral with the same data as in the preceding example, with the exception of the accelerating voltage, which was taken equal to $V_0 = 3000$ v instead of 500 v. The voltage across the surface of the spiral was taken to be somewhat smaller than the optimal voltage (320 v) and is equal to $2V_f = 260$ v.

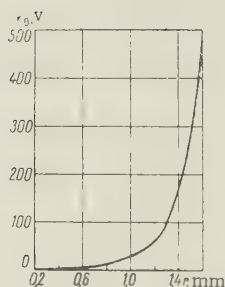


Fig. 7.

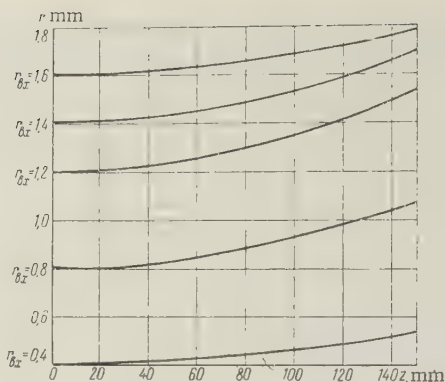


Fig. 8.

Fig. 7. The dependence of the accelerating potential on the radius, $V_0 = f(r)$ for a special electron gun which makes the trajectories of all the electrons in the beam parallel to the axis of the tube ($2V_f = 202$ v, $L = 3.6$ mm).

Fig. 8. The electron trajectories $r = f(z)$ in the beam for different entrance radii for the case $z_v \gg l_{sp}$.

Figure 8 shows that the radial deflection of the electrons of the beam is small, and that electron trajectories are approximately parallel to one another. Thus, it is also possible to obtain proper focusing of a solid beam by means of a periodic electrostatic field with an ordinary electrostatic gun which produces a stream of electrons at the same velocity. This is a new result, since up to now, in accordance with article [7], this type of focusing has been considered impossible.

CONCLUSION

This article gave a theoretical analysis of the focusing of solid and hollow electron beams with periodic electrostatic fields. An examination of the focusing of a hollow beam gave the following results:

1. A method of computing the electron trajectories in a biperiodic focusing device was developed. This made it possible to determine the configuration of the beam in this device. Consequently, it is possible to determine the probability that electrons will strike the bifilar spiral of the tube. It is shown that the electron beam in the focusing device is approximately laminar.

2. A derivation is given for the optimal focusing condition which must be satisfied for the boundary trajectories of the beam to remain approximately parallel to the axis of the tube. It is shown that the optimal voltages for the outer and inner boundary trajectories of the beam are practically identical and that therefore there is no need to use inner and outer bifilar spirals with a different pitch or to supply them with different focusing voltages.

3. It is established that a periodic electrostatic field with a definite length along the axis can serve as a transformer of the cross section of a hollow beam, making it possible to reduce, increase or keep constant the cross section of the beam at the exit from the periodic field relative to its size at the entrance. At the entrance and output from the transformer, the electrons move approximately parallel to the axis of the tube.

4. As a result of an examination of the focusing of a solid beam in a periodic electrostatic field, a method is given for computing the electron trajectories in the beam. A derivation is given for the condition for small undulation of the trajectory at a given radius. The change of the accelerating potential with radius which would be required in a special electron gun producing an electron beam parallel to the axis is determined. It is shown that it is also possible to focus a solid beam by means of a periodic electrostatic field by using an ordinary electron gun producing a stream of electrons at the same velocity.

REFERENCES

1. K.K.N. Chang, Biperiodic electrostatic focusing for high-density electron beams, *Proc. I.R.E.*, 1957, 45, 11, 1522.
2. A.L. Ignitskiy, Design of Periodic Electrostatic Fields Due to Bifilar Helices in Traveling-Wave Tubes, *Radiotekhnika i Elektronika*, 1961, 6, 4, 613.
3. K.K.N. Chang, Confined electron flow in periodic electrostatic fields of very short periods, *Proc. I.R.E.*, 1957, 45, 1, 66.
4. N.V. McLaughlin, Theory and applications of Mathieu functions, transl. under edit. N.N. Denisyuk, IL. 1953.
5. A.L. Igritskiy, Calculation of electron trajectories in a periodic focusing device of a traveling-wave tube. *Radiotekhnika i Elektronika*, 1960, 5, 2, 255.
6. K.K.N. Chang, An electrostatically focused traveling-wave-tube amplifier, *RCA Rev.*, 1958, 19, 1, 86.
7. P.K. Tien, Focusing of a long cylindrical electron stream by means of periodic electrostatic fields, *J. Appl. Phys.*, 1954, 25, 10, 1281.

Received April 4, 1960

GEOMETRIC PARAMETERS OF ELECTRON BEAMS IN IONIC FOCUSING

L.E. Bakhrakh, Yu.D. Zharkov

A simplified method is used to consider the problem of ionic focusing of electron beams having solid cylindrical, ribbon and hollow cylindrical shapes. The results, which are identical in certain cases with the results of a more rigorous analysis, make it possible to compute the necessary geometric parameters of electron beams in ionic focusing

INTRODUCTION

Modern electronic instruments are making increasing use of sharply limited electron beams with a high current density and a great length. Among the nonmagnetic methods of focusing electron beams, the so-called method of ionic focusing presents a certain interest.

It is known that an electron beam passing through gas at a pressure of 10^{-2} to 10^{-3} mm Hg can be focused under the action of the space charge of the positive ions. The qualitative explanation of this phenomenon of ionic focusing is that the electrons of the beam ionize the gas on their path. The ions formed in ionization accumulate on the path of the beam, making a sheeting in the form of a positive space charge, while the secondary electrons disappear rapidly from the beam. In addition to neutralizing the negative space charge of the electrons, the positive ions may also exert an appreciable focusing force. This makes it possible to obtain very long electron beams which are resistant to external perturbations.

The equilibrium stage is the stage in which the number of ions formed per unit time is equal to the number of ions lost from the beam per unit time under the action of the space charge field. This condition is normally expressed by the so-called steady-state equation.

The most rigorous quantitative investigation of ionic focusing of a cylindrical electron beam was conducted in the work of Davidov and Braginskiy [1]. They show in particular that the thermal velocity of electrons plays an extremely important role in ionic focusing, since

the transverse components of the thermal velocity are of the same order as the focusing potential difference between the axis and the edge of the beam. In demonstrating this, they assume the law of adiabatic expansion, according to which the temperature of an electron gas changes when the diameter of the electron beam changes. Ionic focusing of electron beams of rectangular or ring-shaped cross sections was not studied from this point of view.

The present article represents an attempt to examine ionic focusing of electron beams of cylindrical, rectangular, and ring-shaped cross sections on the basis of a simplified method.

1. BASIC EQUATIONS

Usually the question of the pulsation of the edges of the focus electron beam is solved by considering the equation of motion of a boundary electron of the beam. For an ionically focused cylindrical beam of radius r moving along the z -axis, the equation of the radial motion of the electron on the effective boundary can be written in the form

$$m \frac{d^2 r}{dt^2} = f_+ - f_- - p_1,$$

where f_+ is the focusing force of the positive ion; f_- is the defocusing force of the electron space charge; p_1 is the force produced by the transverse thermal velocities of the electron; and m is the electron mass.

We shall assume that the density of the space charge is constant over the cross section of the beam and that the wavelength of the axial variation on the surface of the beam is large relative to its diameter. On the basis of these assumptions, f_+ and f_- can be determined from Gauss's theorem.

Let us also assume that the pressure of the electron gas can be described by the equation $p = RT/V$, where T is the absolute temperature, V is the molar volume, and R is the universal gas constant. Then the force acting on the beam's surface S , enclosing a volume V , will be

$$F = \frac{RT}{V} S.$$

Reducing this force to the force exerted on one electron in this volume, we obtain

$$p_1 = \frac{RT}{NV} S = \frac{2kT}{r},$$

where k is Boltzmann's constant.

For the equation of state of the electron gas, let us take the adiabatic equation $TV^{k-1} = \text{const}$. Then, if we assume that the expansion of the beam against the forces of the field takes place only in the transverse direction, we can reduce this equation to the form

$$\frac{kT}{en} = \text{const},$$

where e is the charge of an electron, and n is the electron concentration. From this it follows that

$$\varphi_T r^2 = \varphi_{TK} r_K^2 = \text{const},$$

where $\varphi_T = kT/e$; $\varphi_{TK} = kT_K/e$; T_K is the cathode temperature, and r_K is the radius characterizing the position of the electron on the cathode.

If we consider the above equations and assume that the axial velocity of electrons v , is constant, we can transform the equation of motion of the boundary electron to the form

$$\frac{d^2 r}{dz^2} = - \frac{\pi \rho_+ r}{U} - \frac{I}{rvU} - \frac{\varphi_{TK} r_K^2}{Ur^3}, \quad (1)$$

where ρ_+ is the average density of the ions in the given cross section of the beam, I is the

current of the beam, and U is the accelerating voltage.

Equation (1) is completely identical with the differential equation for the effective radius of the beam in ionic focusing obtained by the more rigorous method of [1] on the basis of the simultaneous solution of Poisson's equation, the equation of continuity, the equation of state, and the kinetic equation.

Likewise it is possible to write the differential equation for a boundary of electrons on a ribbon-shaped electron beam of width $2y$:

$$\frac{d^2 y}{dz^2} = -\frac{2\pi\rho_+ y}{U} + \pi \frac{I}{vU} + \frac{\varphi_{TK} y_K}{2Uy^2}, \quad (2)$$

where y_K is half the width of the cathode.

It should be noted that Eq. (2) is again completely identical with the differential equation for the effective width of a ribbon-shaped beam which can be obtained on the basis of the method of Davidov and Braginskiy.

Our method can be extended to the case of ionic focusing of hollow cylindrical beams. The equation of motion of electrons on the outer boundary of the beam in this case will be

$$\frac{d^2 r}{dz^2} = -\pi\rho_+ \frac{r^2 - r_a^2}{U r} + \frac{I}{U v r} + \frac{\varphi_{TK} (r_K^2 - r_a^2)}{U (r - r_a) (r^2 - r_a^2)}, \quad (3)$$

where r_a is the radius of the internal boundary of the beam.

By solving any of Eqs. (1), (2) and (3) simultaneously with a steady-state equation which characterized the equilibrium state of the bunch, it is possible to study the pulsation of the boundaries of the electron beams for their respective cases.

For the cylindrical and ribbon-shaped electron beams, the steady-state equation will be respectively [1], [2]

$$\rho_+ = \frac{\sqrt{\frac{m_+}{2e}} s P v \rho_-}{\sqrt{\pi (\rho_+ - \rho_-)}},$$

$$\rho_+ = \frac{\sqrt{\pi} \sqrt{\frac{m_+}{2e}} s P v \rho_-}{2 \sqrt{2} \sqrt{\pi (\rho_+ - \rho_-)}},$$

where ρ_- is the electron density, ρ_+ is the ionic density, s is the specific ionization, P is the gas pressure, m_+ is the mass of an ion.

For a hollow cylindrical beam, the steady-state condition can be written approximately in the form

$$\pi (r^2 - r_a^2) s P \rho_- v = \rho_+ \bar{v}_+ 2\pi r,$$

where \bar{v}_+ is the average radial velocity with which the ions emerge from the beam. This velocity is given by the equation

$$\bar{v}_+ = \frac{1}{2} \sqrt{\frac{2e}{m_+} (\varphi_a - \varphi_b)} =$$

$$= \frac{1}{2} \sqrt{\frac{2e}{m_+} \left[\pi (\rho_+ - \rho_-) (r^2 - r_a^2) + 2\pi (\rho_+ - \rho_-) r_a^2 \ln \frac{r_a}{r} \right]},$$

where φ_a and φ_b are the values of the potential at the inner and outer boundaries of the beam. Hence, we obtain the steady-state equation for the case of a hollow cylindrical beam

$$\rho_+ = \sqrt{\frac{m_+}{2e}} \frac{s P v \rho_-}{\sqrt{\pi (\rho_+ - \rho_-)}} \frac{r^2 - r_a^2}{\sqrt{r^2 - r_a^2 + 2r_a^2 \ln \frac{r_a}{r}}}.$$

2. PULSATION OF THE BOUNDARIES OF A BEAM

We shall look for a solution to the equation of motion in the form

$$\begin{aligned} r &= r_m (1 + \delta), \\ y &= y_m (1 + \delta), \\ r - r_a &= (r_m - r_a) (1 + \delta), \end{aligned}$$

where r_m and y_m are the equilibrium values of the radius and the half-thickness of the beam, being determined from the condition

$$\frac{d^2 r}{dz^2} = 0 \quad \text{or} \quad \frac{d^2 y}{dz^2} = 0,$$

where $\delta \ll 1$ is a quantity characterizing the undulation of the boundaries of the beam.

Under the assumptions we have made, all the differential equations are reduced to linear differential equations of the form

$$\frac{d^2 \delta}{dz^2} + \omega^2 \delta = 0, \quad (4)$$

where the quantity $\omega(P, U, T, s, I)$ determines the wavelength of the boundaries of the beam λ . For the special cases $(\rho_+ - \rho_-)/\rho \gg 1$ and $(\rho_+ - \rho_-)/\rho \ll 1$, the wavelengths of the pulsations are given by the following equations (see Table).

$\frac{\rho_+ - \rho_-}{\rho_-} \gg 1$	$\frac{\rho_+ - \rho_-}{\rho_-} \ll 1$	Remarks
$\lambda = \sqrt{\frac{3}{2}} \pi \left(\frac{\sqrt{\frac{2e}{m_+}} U \Phi_{TK}^{1/2} r_K}{I s P} \right)^{1/2}$	$\lambda = \frac{\pi}{s P} \sqrt{\frac{m}{m_+}}$	Cylindrical beam
$\lambda = \frac{2^{3/2} \pi^{1/2} \left(\frac{2e}{m_+} \right)^{1/4} U^{1/2} \Phi_{TK}^{1/2} y_K^{1/2}}{\left(\frac{7}{3} \right)^{1/2} I^{3/2} s^{3/2} P^{1/2}}$	$\lambda = \frac{4}{\sqrt{3}} \frac{1}{s P} \sqrt{\frac{m}{m_+}}$	Ribbon-shaped beam
$\lambda \simeq \frac{2\pi \sqrt{3}}{\sqrt{7}} \frac{U^{1/2} r_a}{\Phi_{TK}^{1/2}} \left(\frac{r_m}{r_a} - 1 \right)$	$\lambda \simeq \frac{2\pi}{s P} \sqrt{\frac{m}{m_+}} \Phi \left(\frac{r_m}{r_a} \right)$	Hollow cylindrical beam ($r_K = r_m$)

$$\text{Note: } \Phi = \frac{(x^2 - 1)^2}{x^3 (x^2 - 1 - 2 \ln x)} \left[\frac{3}{x} (x^2 + 1) + \frac{1}{x + 1} \right], \quad x = \frac{r_m}{r_a}.$$

The solution of Equation (4) will be

$$\delta = \delta(0) \cos \omega z + \left(\frac{d\delta}{dz} \right)_0 \sin \omega z, \quad (5)$$

where $\delta(0)$ and $(d\delta/dz)_0$ correspond to the initial conditions for $z = 0$.

We must note that our study of ionic focusing has been made without considering the magnetic forces of the moving electron beam, since it is easily shown that for nonrelativistic velocities ($c/v > 10$) the focusing force produced by the excess density of the ionic space charge is at least one order of magnitude greater than the force of electromagnetic attraction of electrons in the case of real beams.

By studying Equation (5) it is possible to determine the dependence of the geometric parameters of ion-focused beams (the equilibrium radius, the wavelength of the pulsation, etc.) on the quantities determining the condition of ionic focusing (the gas pressure, the

electron velocity, the beam current, etc.). The general nature of the resulting relationships turns out to be identical for electron beams of different configuration.

Figure 1 shows a curve of the change of the equilibrium width of an ion-focused ribbon-shaped beam of electrons as a function of pressure. The decrease of the equilibrium width with increasing pressure is related to the increase in the focusing force of the positive ions, which is in turn caused by the increase in ionization. Figure 2 gives an analogous curve for a hollow cylindrical electron beam.

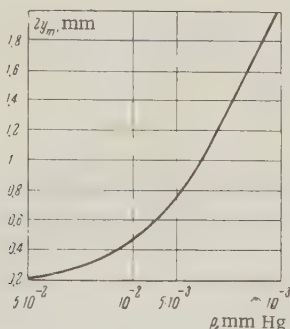


Fig. 1.

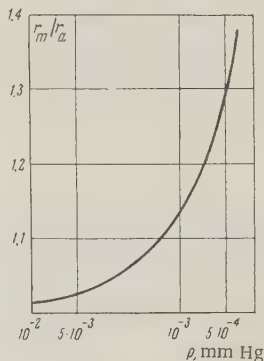


Fig. 2.

Fig. 1. Curve of the change of the equilibrium width of ion-focused, ribbon-shaped electron beam as a function of pressure ($I = 30$ ma, $U = 500$ v, $2y_k = 1.2$ mm)

Fig. 2. Curve of the change in the equilibrium radius of an ion-focused hollow cylindrical electron beam as a function of pressure ($I = 170$ ma, $U = 300$ v, $r_m = r_k$)

Figure 3 shows the equilibrium width of a ribbon-shaped beam as a function of the accelerating voltage. The increase of the equilibrium width of the beam with increasing accelerating voltage is caused chiefly by the decrease of the probability of ionization, with the resulting decrease of the focusing forces and expansion of the beam. An analogous curve is given for a hollow cylindrical beam in Figure 4.

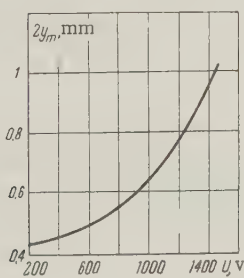


Fig. 3.

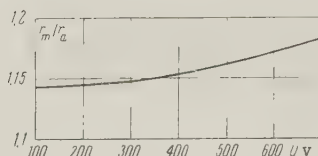


Fig. 4.

Fig. 3. The equilibrium width of a ribbon-shaped beam of electrons as a function of the accelerated voltage ($I = 40$ ma, $p = 10^{-2}$ mm Hg, $2y_k = 1$ mm)

Fig. 4. The equilibrium radius of a hollow cylindrical electron beam as a function of the accelerating voltage ($I = 170$ ma, $p = 10^{-3}$ mm Hg, $r_m = r_k$)

Figure 5 illustrates the dependence of the wavelength of the pulsations on pressure for a cylindrical electron beam. From this figure, it is seen that the wavelength decreases with increasing gas pressure. This is the result of the increase in focusing force in the beam.

Figure 6 shows the nature of the change of the wavelength of the pulsations of the outer

boundary of a hollow cylindrical beam with voltage.

Figures 7 and 8 show the amplitude of the pulsations δ_{\max} as a function of pressure and the accelerating voltage for the case of an ion-focused ribbon-shaped beam. For these curves,

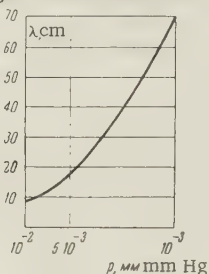


Fig. 5.

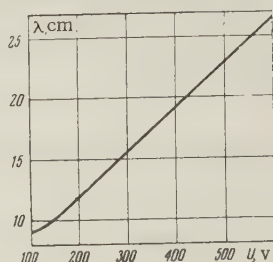


Fig. 6.

Figure 5. The wavelength of the pulsations as a function of the pressure for a cylindrical electron beam ($U = 800$ v, $r_m = r_k$)

Figure 6. The wavelength of the pulsations of the outer boundary of a hollow beam of electrons as a function of voltage ($I = 170$ ma, $q = 10^{-3}$ mm Hg, $r_m = r_k$)

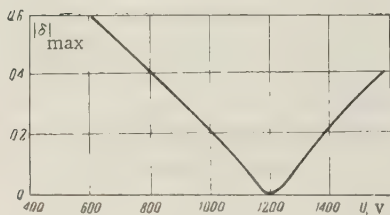


Fig. 7.

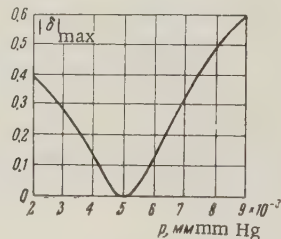


Fig. 8.

Fig. 7. The amplitude of the pulsations as a function of the accelerating voltage in ionic focusing of a ribbon-shaped electron beam ($I = 40$ ma, $p = 10^{-2}$ mm Hg, $\gamma = 0$, $y_k = 0.4$ mm)

Fig. 8. The amplitude of the pulsations as a function of pressure in ionic focusing of a ribbon-shaped electron beam ($I = 40$ ma, $U = 500$ v, $\gamma = 0$, $y_k = 0.4$ mm)

it is assumed that the beam is injected parallel to the axis ($\gamma = 0$). From Figures 7 and 8, it follows that there exists optimal pressures and voltages at which the pulsations are minimal. These optimal values of the pressure and the voltage correspond to the case where the equilibrium width of the beam is equal to the width of the cathode.

REFERENCES

1. B.I. Davidov, S.I. Braginskiy. The theory of the gas concentration of electron beams, Collection of articles dedicated to the 70th Anniversary of Academician A. F. Ioffe. AN SSSR, 1950, 72-91
2. M.M. Bredov. Automatic compensation for the space charge electron beams, Collection of articles dedicated to the Seventieth Anniversary of Academician A. F. Ioffe, AN SSSR, 1950, 155-172.

Received June 18, 1960

THE PROBLEM OF CHANGING THE SIGNAL-TO-NOISE RATIO PRODUCED BY CAMERA TUBES IN A TV SYSTEM

S. B. Gurevich

The reduction of the signal-to-noise ratio from the input to the output of a TV camera tube is considered in general form. The possibilities of raising the sensitivity by connecting the camera tube with an image intensifier are examined.

INTRODUCTION

One of the most important directions of the future development of television lies in raising the sensitivity of TV cameras. The problem of achieving a sensitivity which is close to the limiting sensitivity and which is determined almost entirely by the fluctuations in the quanta of radiation is not at all fictitious.

To solve such a problem, it is necessary to understand the characteristic features of the transformation of the signal and the noise between the input of the camera tube and its output. Below we shall consider in general form the changes in the signal-to-noise ratio occurring in different sections of camera tubes, and we consider the possibility of using an image intensifier for one of the links for the purpose of drawing considerably nearer to the limiting sensitivity.

1. THE SENSITIVITY AND SIGNAL-TO-NOISE RATIO OF A TV SYSTEM

Until recently, the sensitivity of a TV system was understood to be a quantity which was the inverse of the illumination of the object required for a certain given quality of the image in the receiver. This definition is, in fact, bound to one standard scanning system (used in TV broadcasting) and it cannot be extended to different television systems without the introduction of additional systems.

A more general definition of the sensitivity of cameras is provided by a quantity which is the reciprocal of the minimum energy radiation for the surface of the object for which the television system can transmit a given amount of information about the object. This amount of information is determined by the number η of distinguishable quanta (thresholds) of the energy, the number ν of separately distinguishable elements of the object, and the number κ of states of the objects.

In accordance with this definition, the following equation for the sensitivity G of TV equipment can be obtained [1]:

$$G = g \frac{1}{\frac{\eta}{2} (\eta + 1) \nu \kappa} \quad (1)$$

where

$$g = \frac{\varepsilon_c^2}{\gamma \varepsilon_{ob}}$$

is the specific sensitivity. This specific sensitivity is the reciprocal of the energy which one

element of the object would have to radiate for one level of radiant energy to be distinguished in a real system. In the equation for the specific sensitivity, ε_{ob} is the corresponding energy which would be required in an ideal system, while γ is a coefficient characterizing the loss of radiant energy in optical projection of the image of the object on the photosensitive layer. The quantity ξ_c characterizes the change of the signal-to-noise ratio between the input and the output:

$$\xi_c = \frac{\psi_{outc}}{\psi_{in}},$$

where ψ_{in} is the signal-to-noise ratio at the input of the camera (on the photosensitive surface of the camera tube), and ψ_{out} is the same ratio at the output of the system.

In regard to the camera tube, let us express g in the form

$$g = \frac{\xi^2}{\varepsilon_0}. \quad (2)$$

Here $\varepsilon_0 = \gamma \varepsilon_{ob}$ is the radiation energy falling on one element of the image of the photosensitive layer, and $\xi = \psi_{out}/\psi_{in}$, where ψ_{out} is the signal-to-noise ratio at the output of the camera tube rather than at the output of the system.

From Eqs. (1) and (2), it is seen that the change of the signal-to-noise ratio which is caused by the camera tube and which is characterized by the coefficient ξ is directly connected with the sensitivity of the system and the camera tube.

The definition of the signal-to-noise ratio in a camera tube performing a series of transformations requires separate computation of the change of the magnitude of the signals and effective value of the noise. In the conversion of radiant energy into electric energy and of electric energy back to radiant energy which can occur several times, the information has different carriers. Thus, for the above-mentioned computations of the signal and the noise, it is necessary to define more closely the concept of the signal and the noise. In the majority of works devoted to noise in television systems and other communications systems, the signal* (which we denote by S) is understood to be a current changing with time or the corresponding voltage: $S = i(t)$ or $S = U(t)$. It is obvious that the fluctuations of these quantities have the same dimensions. This concept of the signal is, in particular, the result of the fact that in Shannon's circuit of a general system for communications [2] a source of noise is contained only in the devices which transmit the signal after the message has been converted to electrical energy. It is obvious that for the determination of the signal-to-noise ratio in this case it is sufficient to use some electrical quantity for the signal (for instance, the current or the voltage).

In contrast to Shannon's system, in the more general case it is necessary to consider a system in which noise is generated together with the signal by the actual source before conversion to electric quantities. Moreover, it is both possible and advisable [4, 5, 6] to have several conversions of the radiation into electron beams and back again, during which both the signal and the noise will change. Consequently, it is necessary to establish a concept of the signal which will not depend on the physical properties of the carrier of information and which will make it possible to retain the proportionality of the signal to the radiant energy required to transmit one unit of information from the object (one nonzero gradation of energy in an element which is scanned once). A suitable concept of this type is the number of particles μ conveying information. In the case of undistorted information, this number is proportional to the number of quanta required to televise a given number of gradations, elements, and states of the object. If we denote the signal magnitude by S , we can write $S = \mu$. In this case, the square of the effective value of the noise N^2 can be represented by the variance of the number of particles; $N^2 = \overline{\Delta\mu^2}$. In the case where the shot effect is present and the Poisson or Gauss distribution holds, the variance of the number of particles is equal to $\bar{\mu}$, and thus

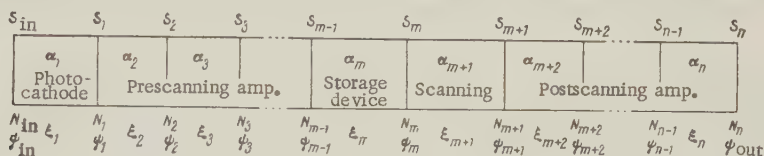
$$N^2 = \overline{\Delta\mu^2} = \bar{\mu} = \bar{S}.$$

Let us now evaluate ξ for TV camera tubes.

*A. A. Kharkevich [3] defines the signal as the conveyor of time-dependent information.

2. THE SIGNAL-TO-NOISE RATIO IN TV CAMERA TUBES

For the sake of convenience, we shall break up the camera device into a series of stages (see figure), in each of which the signal-to-noise ratio changes. This change occurs as a result of changes in the signal and the noise. The signals will change in accordance with the



amplification properties of the stages, each of which has its own amplification coefficient. The changes in the noise can have a dual nature. In the case where the stage is an energy converter independently of whether it amplifies, attenuates or does not change the signal, the square of the effective value of the noise in this stage is quantitatively equal to the sum of the squares of the noise from the preceding stage multiplied by the amplification coefficient α and the noise formed in the given stage. This quantity is equal to the square root of the value of the signal at the output of this stage.

In some stages, noise may also appear from extraneous sources not connected with the value of the signal, for instance from the scanning beam.

The camera receives noise together with the radiation from the beam. This noise is attenuated when the radiation is converted into photoelectric currents (with a quantum efficiency less than unity) and is then sometimes amplified before reaching the storage device (prescanning amplification). The storage device accumulates part or all of the charge, and correspondingly the noise and the signal are preserved or (depending on the efficiency of storage) are reduced.

In scanning, all or part of the stored charge travels on as the useful signal. Simultaneously with the scanning or after the scanning, post-scanning amplification through secondary emission may take place. This will magnify both the signal and the noise. Finally, after emerging from the camera tube, both the signal and the noise are amplified and attenuated in various amplifier stages.

The noise arising in each successive stage undergoes the same changes, starting with the stage in which it originates. We shall consider the TV camera tube as having n amplifying stages with amplification coefficients $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_i, \dots, \alpha_n$, some of which are greater than unity, and the remainder of which are smaller than unity.

Let the coefficients $\xi_1, \xi_2, \xi_3, \dots, \xi_i, \dots, \xi_n$ denote the change of the signal-to-noise ratio in the corresponding stages, i.e.,

$$g_{ij}^{\pm} = \frac{\psi_i}{\psi_{i-1}}. \quad (3)$$

It can be easily shown that these coefficients are always smaller than unity, i.e., that the sequence of numbers $\psi_{in}, \psi_1, \psi_2, \dots, \psi_i, \dots, \psi_n$, formed from the signal-to-noise ratios at the output of each successive stage, satisfy the following condition:

$$\psi_{in} > \psi_1 > \psi_2 > \dots > \psi_i > \dots > \psi_n \quad (4)$$

($\psi_{in} = S_{in}/N_{in}$ is the input-to-noise ratio characterizing the given radiant energy).

Let us consider the changes of the signal-to-noise ratio occurring in those converting stages of the camera tube in which the noise depends only on the magnitude of the signal*.

*In real stages which convert radiant energy into photoelectric current, there is actually, strictly speaking, additional noise which does not depend on the signal (noise resulting from fluctuations in the dark current electrons). However, it turns out that in many cases this noise can be neglected [7].

We shall suppose that there are m such stages.

At the output of the first stage, the square of the noise N_1^2 will be characterized quantitatively by the sum of the squares of the amplified noise at the input $N_{in}^2 \alpha_1^2$ and the value of the noise formed in the first stage. The noise formed in this stage is equal to the signal of the output of this stage, i.e., $S_1 = S_{in} \alpha_1$. Consequently,

$$N_1^2 = N_{in}^2 \alpha_1^2 + S_{in} \alpha_1 = \alpha_1 (\alpha_1 + 1) S_{in}, \quad (5)$$

since $N_{in}^2 = \bar{S}_{in}$.

At the output of the second stage,

$$N_2^2 = N_1^2 \alpha_2^2 + S_2 = \alpha_2^2 (\alpha_1^2 + \alpha_1) S_{in} + \alpha_1 \alpha_2 S_{in} = \alpha_1 \alpha_2 [\alpha_2 (\alpha_1 + 1) + 1] S_{in}.$$

At the output of the third stage,

$$\begin{aligned} N_3^2 &= N_2^2 \alpha_3^2 + S_3 = \alpha_3^2 \alpha_1 \alpha_2 [\alpha_2 (\alpha_1 + 1) + 1] S_{in} + \alpha_1 \alpha_2 \alpha_3 S_{in} = \\ &= \alpha_1 \alpha_2 \alpha_3 [\alpha_3 (\alpha_2 (\alpha_1 + 1) + 1) + 1] S_{in} = \alpha_1 \alpha_2 \alpha_3 (\alpha_1 \alpha_2 \alpha_3 + \alpha_2 \alpha_3 + \alpha_3 + 1) S_{in} \end{aligned}$$

and finally at the output of the m -th stage

$$N_m^2 = N_{m-1}^2 \alpha_m^2 + S_m = \alpha_1 \alpha_2 \alpha_3 \dots \alpha_m (\alpha_1 \alpha_2 \alpha_3 \dots \alpha_m + \alpha_2 \alpha_3 \dots \alpha_m + \dots + \alpha_m + 1) S_{in}.$$

We set $\alpha_1 \alpha_2 \alpha_3 \dots \alpha_i = M_i$. Then

$$\begin{aligned} \alpha_1 \alpha_2 \alpha_3 \dots \alpha_m + \alpha_2 \alpha_3 \dots \alpha_m + \alpha_3 \dots \alpha_m + \dots + \alpha_m + 1 &= \frac{M_m}{1} + \frac{M_m}{M_1} + \frac{M_m}{M_2} + \\ + \frac{M_m}{M_3} + \dots + \frac{M_m}{M_{m-1}} + \frac{M_m}{M_m} &= M_m \left(1 + \frac{1}{M_1} + \frac{1}{M_2} + \dots + \frac{1}{M_{m-1}} + \frac{1}{M_m} \right) = \\ &= M_m \left(\sum_{j=1}^m \frac{1}{M_j} + 1 \right). \end{aligned}$$

In this case

$$N_i^2 = M_i^2 \left(\sum_{j=1}^i \frac{1}{M_j} + 1 \right) S_{in}. \quad (6)$$

The value of the signal-to-noise ratio of the i -th stage is

$$\psi_i = \frac{S_i}{N_i} = \frac{M_i S_{in}}{M_i \sqrt{\left(\sum_{j=1}^i \frac{1}{M_j} + 1 \right) S_{in}}} = \sqrt{\frac{S_{in}}{\sum_{j=1}^i \frac{1}{M_j} + 1}}. \quad (7)$$

The coefficient characterizing the change of the signal-to-noise ratio in the i -th stage is

$$\eta_i = \frac{\psi_i}{\psi_{i-1}} = \sqrt{\frac{\sum_{j=1}^{i-1} \frac{1}{M_j} + 1}{\sum_{j=1}^i \frac{1}{M_j} + 1}} = \sqrt{\frac{M_i \left(\sum_{j=1}^{i-1} \frac{1}{M_j} + 1 \right)}{M_i \left(\sum_{j=1}^i \frac{1}{M_j} + 1 \right) + 1}}, \quad (8)$$

since

$$\sum_{j=1}^i \frac{1}{M_j} = \sum_{j=1}^{i-1} \frac{1}{M_j} + \frac{1}{M_i}.$$

It is obvious that ξ_i can change within the limits

$$0 \leq \xi_i < 1$$

and that the inequality

$$\psi_{\text{in}} > \psi_1 > \psi_2 > \dots > \psi_i > \dots > \psi_m.$$

is correct

From Eq. (8) it is seen that the value of the signal-to-noise ratio in the given stage will remain practically constant ($\xi = 1$) if

$$M_i \left(\sum_{j=1}^{i-1} \frac{1}{M_j} + 1 \right) = \alpha_1 \alpha_2 \alpha_3 \dots \alpha_i + \alpha_2 \alpha_3 \dots \alpha_i + \dots + \alpha_i \gg 1. \quad (9)$$

This can occur in the case where, for instance, the amplification coefficient of the i -th stage $\alpha_i \gg 1$ (independently of the value of the remaining coefficients), or in the case where α_i is small while the other coefficients are much larger than unity. If the signal-to-noise ratio is to change very little throughout all m stages, the product of the factors $\xi_1 \xi_2 \xi_3 \dots \xi_m = \psi_m / \psi_{\text{in}}$ must be very close to unity, i.e.

$$\frac{\psi_m}{\psi_{\text{in}}} = \sqrt{\frac{1}{\sum_{j=1}^m \frac{1}{M_j} + 1}} = 1 - \frac{\delta}{2},$$

where δ is a small quantity satisfying the condition $\delta \ll 1$. Consequently

$$\frac{1}{\alpha_1} + \frac{1}{\alpha_1 \alpha_2} + \frac{1}{\alpha_1 \alpha_2 \alpha_3} + \dots + \frac{1}{\alpha_1 \alpha_2 \alpha_3 \dots \alpha_m} < \delta. \quad (10)$$

Thus, for the signal-to-noise ratio to change little between the input and the output of the camera, it is necessary that each term in inequality (10) be sufficiently small. This means that all of the denominators be much greater than unity. For this, it is necessary first of all that the amplification coefficient of the first stage satisfy the condition $\alpha_1 = 1$. If one of the succeeding factors (for instance, α_2) is smaller than unity, the preceding factors (in the given case) must be sufficiently large for the product of these products and the succeeding factors (in the given case α_1) to still satisfy the condition $\alpha_1 \alpha_2 \gg 1$. If this is not the case, inequality (10) (which has the form $\frac{1}{\alpha_1} + \frac{1}{\alpha_1 \alpha_2} < \delta \ll 1$ for the given example of two stages) will

not be satisfied. As a result, the signal-to-noise ratio will be considerably reduced in the second stage and will subsequently remain almost the same or will be still further lowered.

Thus, to approach a camera device with ideal sensitivity, it is first of all necessary that the amplification coefficients in the first stage be sufficiently high.

The equations we have obtained characterize the change of the signal-to-noise ratio in stages in which the level of noise is almost entirely determined by the input signal and in which there are no extraneous sources of noise the magnitude of which does not depend directly on the magnitude of the input signal. As we have already pointed out, scanning may be a source of such noise in camera tubes (in scanning, the beam current is not connected with changes in the value of the input signal).

Let us assume that the scanning by electron beams takes place in the $m+1$ -th stage. At the output of this stage, the total number of electrons will be a linear combination of the number of electrons forming the total signal and the scanning beam. Consequently, the number of electrons corresponding to a given element will be given by

$$S' = \alpha_{m+1} S_m + \beta S_n = \alpha_{m+1} M_m S_{\text{in}} + \beta S_n, \quad (11)$$

where α_{m+1} and β are coefficients depending on the properties of the scanning mechanism and of the method of forming the output signal in different types of camera tubes. One of these

coefficients will assume a negative value, as for instance the case of a superorthicon, in which a beam of electrons consisting of the difference between the beam current and the signal current is subject to further amplification.

The noise in the scanning stages can now be described as follows:

$$N_{m+1}^2 = \alpha_{m+1}^2 N_m^2 + \alpha_{m+1} S_m + \beta^2 N_n^2 + \beta S_n = \alpha_{m+1}^2 M_m^2 \left(1 + \sum_{j=1}^m \frac{1}{M_j}\right) S_{in} + \\ + \alpha_{m+1} M_m S_{BX} + \beta^2 \left(1 + \frac{1}{\beta}\right) S_n = M_{m+1}^2 \left(1 + \sum_{j=1}^{m+1} \frac{1}{M_j}\right) S_{in} + \beta \left(1 + \frac{1}{\beta}\right) S_n.$$

In the following stages with amplification coefficients α_{m+2} , α_{m+3} , ..., α_n , the changes of the two components of noise will be determined by the same factors. It is easily shown that the square of the effective value of the noise at the output of the n -th stage is given by the equation

$$N_n^2 = M_n^2 \left(1 + \sum_{j=1}^n \frac{1}{M_j}\right) S_{in} + \frac{M_n^2 \beta^2}{M_m^2 \alpha_{m+1}^2} \left(1 + \frac{M_m \alpha_{m+1}}{\beta} \sum_{j=m+1}^n \frac{1}{M_j}\right) S_n. \quad (12)$$

Thus, for the signal-to-noise ratio of the n -th stage, we obtain the quantity

$$\psi_n = \sqrt{\frac{S_{in}}{1 + \sum_{j=1}^n \frac{1}{M_j} + \frac{\beta^2 S_n}{M_m^2 \alpha_{m+1}^2 S_{in}} \left(1 + \frac{M_m \alpha_{m+1}}{\beta} \sum_{j=m+1}^n \frac{1}{M_j}\right)}}. \quad (13)$$

By analogy with the above argument, we note that the change of the signal-to-noise ratio between the input and the output of the camera tube will be small if

$$\sum_{j=1}^n \frac{1}{M_j} + \frac{\beta^2 S_n}{M_m^2 \alpha_{m+1}^2 S_{in}} \left(1 + \frac{M_m \alpha_{m+1}}{\beta} \sum_{j=m+1}^n \frac{1}{M_j}\right) < \delta, \quad (14)$$

where $\delta \ll 1$. For this condition to be satisfied, it is necessary that the inequality

$$\beta^2 S_n \ll M_m^2 \alpha_{m+1}^2 S_{in} \text{ и } \beta S_n \ll M_m \alpha_{m+1} S_{in}, \quad (14)$$

be satisfied in addition to the requirements for the coefficients α_1 , α_2 , α_3 , ..., α_n given in condition (10). For this, it is necessary that the scanning amplification, given by the product of factors $M_m = \alpha_1 \alpha_2 \alpha_3, \dots, \alpha_m$, be sufficiently large and that it satisfy the condition

$$M_m \gg \frac{\beta S_n}{\alpha_{m+1} S_{in}}. \quad (15)$$

It is easily seen that condition (4) which characterizes the reduction of the signal-to-noise ratio from stage to stage still holds in the case where Eq. (13) is valid.

In existing tubes, the "ideal" conditions where the ψ of each successive stage differs little from the ψ of the preceding stage do not hold for all the stages. When camera tubes such as the supericonoscope and the vidicon are used, considerable additional noise forms in the stages beyond the camera tube (at the input of the preamplifier). This additional noise is usually so large relative to the noise of the actual camera tube, that a signal-to-noise ratio is basically determined by the ratio of the signal from the camera tube to the noise of the input and the first stage of the preamplifier. In the superorthicon, there is a considerable decrease of the signal-to-noise ratio in the scanning stage, since inequality (15) is not satisfied. It is obvious that the ways of approaching ideal camera devices lie in the attempts to achieve as large an amplification as possible in the first stage (prescanning amplification, the necessity of which is emphasized by G. L. Gerus [4]) for camera tubes of the superorthicon type and in the use of prescanning amplifications plus postscanning amplification through secondary emission in camera tubes of the supericonoscope and vidicon types.

Let us determine how many times the light flux falling on the photosensitive surface of a LI-17 superorthicon is greater than the light flux which would fall on the corresponding photosensitive surface of an ideal camera device, if we are required to transmit $\eta = 20$ gradations of brightness, $v = 400,000$ elements and $\kappa = 25$ frames. With $\xi = 1$ and the corresponding scaling coefficients, we can determine from Eq. (1) the minimum light flux necessary for transmission of the image of the ideal system under the above conditions. This light flux amounts to $\sim 0.65 \cdot 10^{-6}$ lumens.

From the literature [8], it is known that the minimum light flux required to transmit a picture in the case of an LI-17 tube is $0.4 \cdot 10^{-3}$ lumens. Consequently, the required light flux is approximately 600 times greater than the corresponding light flux for an ideal system. From (1), it follows that $\xi = \psi_{\text{out}} / \psi_{\text{in}} = \sqrt{600} \approx 25$, i.e., the signal-to-noise ratio decreases by a factor of approximately 25 between the input and the output of a LI-17 superorthicon camera tube. This decrease of the signal-to-noise ratio takes place in the following section of the camera tube: a) the photocathode; b) the storage section; and c) the scanning section. In the other sections, the signal-to-noise ratio is reduced very little. In actual LI-17 tubes, the coefficient characterizing the photocathode is equal to approximately 1/20. This is the part of the sensitivity loss which can be reduced only a little at the present time by increasing the total sensitivity 2 - 4 times (for instance, by reducing α_1 to 1/5 in multialkaline photocathodes).

On the other hand, it is possible to increase considerably the sensitivity by introducing coefficients α_2, α_3 , etc. which are considerably larger than unity, i.e., by using prescanning amplification. One of the methods of achieving prescanning amplification is the use of an electron - optical amplifier, which can be placed in the same casing as the superorthicon [4, 5] or which can be used separately [6]. In both cases, the total amplification must be considerably greater than 30 (since the total required increase of the light flux is represented by a factor of 600, and the attenuation coefficient at the photocathode is approximately 1/20). The use of a single stage electron - optical amplifier with an amplification coefficient of the order of 30 - 60 in the same casing as the camera will make the photocathode noise of the same order as the scanning noise. The sensitivity of this device will be 1/2 to 2/3 that of a camera device in which the sensitivity is limited only by noise from the photocathode. The other completely satisfactory system is the use of a separate multistage electron - optical amplifier combined with intermediate optical devices. In this case, the amplification coefficient of the amplifier must be much larger than 1000 (in view of the coefficient characterizing the losses in the optical system).

Thus, there are real possibilities of making television camera devices much closer to ideal cameras.

3. COMMENTS

We said above that the storage which was carried out in camera tubes can be represented by an amplifying stage whose amplification coefficient is smaller than unity. The theoretical basis which makes storage possible is connected with the fact that in television the picture element is repeated with a limited velocity, determined by the number of frames per second. The amplification coefficient characterizing the storage stage is equal to unity in the case where all the electrons formed as a result of single or multiple versions of light into photocurrent during the time interval between two scanings of the same element are collected on the given portion of the storage device. We shall attribute the secondary emission occurring on the surface of the storage device to the separate stage for prescanning amplification. The amplification coefficient of the storage stage must not exceed unity, since, if this were the case, there would be an accumulation of charge formed during a time exceeding the time of a single frame.

This would lead to sluggish transmission and, consequently, to a reduction of the amount of transmitted information. Hence, it follows that it is pointless to use several storage devices in the place of a single efficient one with an amplification coefficient close to unity.

In recent years it has been argued that it is necessary to include the luminophores (phosphors) of electron-optical amplifiers, which become luminescent slowly and have long persistence among the storage devices used in camera tubes [9, 10]. It was stated in these articles that the use of these amplifiers would make it possible to increase the output signal-to-noise ratio relative to the input signal-to-noise ratio. In reality, the signal-to-noise ratio can not be increased with respect to preceding stages. It can only be increased by changing

other conditions of transmission, these conditions of transmission holding equally for all the stages, including both the input and the output stages. When it is necessary to increase the signal-to-noise ratio, these transmission conditions can be changed only in a single direction, namely by reducing the number of elements or by reducing the number of scannings of these elements. It is obvious that when the picture is divided into a smaller number of elements (at the input, this imaginary division is made on the basis of the scanning pattern which is used and the dimensions of the scanning beam) or when the number of scannings of these elements is reduced, and individual elements will contain a greater amount of radiation energy than in the case where a large number of elements with a large number of frames per second is transmitted. Thus, these changed conditions will result in a large input signal with a large signal-to-noise ratio. The condition according to which the signal-to-noise ratio between the input and the output can only decrease remains in force. Any assertion to the contrary normally contains an inaccurate definition and sometimes an error in the derivation. In [9], there is an attempt to demonstrate that the signal-to-noise ratio can increase between the input and the output when an electron-optical amplifier with a luminophore is used. In this article, the mean square fluctuation of the current obtained after amplification on a layer consisting of a high-persistence luminophore and a zero-persistence photocathode is determined. The equation for the mean square current is given in the form

$$i_n^2 = 2ei_1 \left[m_1^2 q_2^2 \int_0^\infty \varphi_0^2(t) \varphi_1^2(t) dt + (m_1 q_2^2 + m_1 q_2) \int_0^\infty \varphi_0^2(t) dt \right], \quad (16)$$

where i_1 is the signal current; q_2 is the quantum efficiency of the photocurrent; m_1 is the number of light quanta produced by one electron absorbed by the luminophore; $\varphi_0(t)$ is the frequency characteristic of the receiver of the photocurrent formed after amplification; and $\varphi_1(t)$ is the frequency characteristic of the luminophore.

The error made in Eq. (16) consists of the fact that the frequency characteristic of the second component of the noise is taken to be independent of the frequency characteristic of the phosphor. From a derivation of the equations for the fluctuation of the number of electrons given the Poisson or Gaussian distribution, it follows that in reality the frequency characteristic of both components should be identical. Thus, in the second term of Eq. (16), the factor $\int_0^\infty \varphi_0^2(t) dt$ should be replaced by $\int_0^\infty \varphi_0^2(t) \varphi_1^2(t) dt$. If the factor characterizing the passband is taken outside the brackets, the equation

$$\psi_1^2 = \frac{\psi_0^2}{\int_0^\infty \varphi_1^2(t) dt \left/ \int_0^\infty \varphi_0^2(t) dt + \frac{q_2 + 1}{m_1 q_2} \right.},$$

in which the denominator is smaller than unity [since, from [9], $\int_0^\infty \varphi_1^2(t) dt \ll \int_0^\infty \varphi_0^2(t) dt$, and $(q_2 + 1)/m_1 q_2 \ll 1$] will be replaced by the equation $\psi_1^2 = \frac{\psi_0^2}{1 + \frac{q_2 + 1}{m_1 q_2}}$, in which the denominator

is greater than unity. From this it follows that the input signal-to-noise ratio is still greater, and not smaller, than the output ratio. A similar error is made in [10]. There, in a computation of the power of the shot noise at the output of three stages, the amplification at the phosphor - photocathode layer is given as

$$i_{\text{ш}} = \frac{ei_1 \beta_1^2 \alpha_2^2 \alpha_3^2}{V \theta_1^2 + \theta_2^2 + \theta_3^2} + \frac{ei_2 \beta_2^2 \alpha_3^2}{V \theta_2^2 + \theta_3^2} + \frac{ei_3 \beta_3^2}{\theta_3} + \frac{ei_4}{\tau}, \quad (17)$$

where i_1 is the output current, i_2 is the current after one amplifying stage; i_3 is the current after two amplifying stages; i_4 is the output current after all four amplifying stages; β_1 , β_2 and β_3 are the amplification coefficients in the corresponding stages; θ_1 , θ_2 , and θ_3 are quantities called the effective storage time of the phosphor by the authors of [10]; and τ is the

quantity called the correlation time. From the above equation, it is seen that here noise belonging to different intervals of time is added up at the output. Since the time in the first three terms of Eq. (17) are considerably larger than that of the fourth term, the signal-to-noise ratio is basically determined only by the magnitude at the output current

$\Psi_4 \approx \sqrt{\frac{i_4 \tau}{e}}$. Since $i_4 = \beta_1 \beta_2 \beta_3 i_1$, it seems that there is a gain in the signal-to-noise ratio

of $\sqrt{\beta_1 \beta_2 \beta_3}$ times between the input and the output (the authors of [10] find an increase of $\beta_1 \beta_2 \beta_3$ times. It is obvious that if the transmission time of a single element is taken to be a constant quantity, the signal-to-noise ratio at the output will be smaller than at the input, as does indeed occur.

CONCLUSIONS

1. The sensitivity of TV camera tubes depends both on the factor characterizing the minimum radiation energy sufficient to transmit η gradations, ν elements, and κ states of the object in the limiting case of restrictions imposed only by fluctuations in the radiated quanta, and on the factor ξ which characterizes the change in the signal-to-noise ratio between the input and the output. The value of this second factor depends to a considerable extent on the properties of the camera tube.

2. The signal-to-noise ratio of a camera device transmitting a given number of elements and a given number of states of the scene can only decrease between the input and the output of the camera device. This decrease in the signal-to-noise ratio takes place independently of the persistence of the converters used in the camera device. If the storage is efficient and the coefficient characterizing the storage efficiency is equal to unity, the use of persistent converters will only produce distortions in the information about the scene which is transmitted (aftereffects), without raising the signal-to-noise ratio.

3. A signal-to-noise ratio which is reduced in one of the initial stages cannot then be restored to its initial value whatever the amplification in the succeeding stages. The reduction of the signal-to-noise ratio in a stage whose amplification coefficient is smaller than unity can be prevented only if this stage is preceded by at least one stage with an amplification coefficient much greater than unity.

4. The superorthicon model (LI-17 and LI-201) used in TV work has a sensitivity approximately three orders of magnitude smaller than the limiting sensitivity. The use of electron-optical amplifiers in the same units as the superorthicon or separately, makes it possible to raise the sensitivity of the superorthicon by approximately two orders of magnitude and to approach a sensitivity of the system determined only by fluctuations in the photocurrent.

REFERENCES

1. S.B. Gurevich. Threshold and contrast sensitivity of TV systems with different scanning patterns. Reports presented at the scientific and technical conferences dedicated to Radio Day in Leningrad and Moscow, 1960.
2. K. Shannon. Statistical theory of the transmission of electric signals. Collection of articles on the theory of transmission of electric signals with noise. IL, 1953, pp. 7-87.
3. A.A. Kharkevich. Outline of the General Theory of Communications. GTI, 1955.
4. V.L. Gerus. Electronic processes in TV camera tubes with a two-phase target. Dissertation, 1952.
5. C.A. Morton, J.E. Ruedy, The intensifier orthicon, Conference Proceedings 2nd National Convention on Military Electronics, PGME IRE, 1958, p. 113-117.
6. A.M. Khalfin. Electronic amplification of images. Tekhnika Kino i Televedeniya, 1958, 9, 10.
7. E.I. Zavoytskiy, M.M. Butslov, G.B. Smolkin. The limiting amplification coefficient and the noise of the electron-optical light amplifiers. Doklady AN SSSR 1956, vol. 111, 65, 996.
8. S.B. Gurevich. Physical processes in TV camera tubes. GIFML, 1958, p. 225.
9. P.V. Makovetskiy. The signal-to-noise ratio in electron-noise optical amplifiers, Zh. Tekh. Fiz. 1959, vol. 29, 3, 406.

Received June 10, 1960

MATHEMATICAL FORMULATION OF PROBLEMS IN CORPUSCULAR OPTICS WITH SPACE CHARGE TAKEN INTO ACCOUNT

V. M. Breitman, V. S. Kuznetsov

Articles [1, 2, 3 and 4] give the results of the development of a method of computation for corpuscular optical systems in which the effect of space charge is taken into account by the method of a model [5, 6]. This article contains a mathematical formulation of two types of boundary problems ("direct" and "converse") for a nonrelativistic corpuscular beam whose particles have curvilinear trajectories. This formulation is given in connection with the development and methods of solution of these problems by rapid electronic computers.

INTRODUCTION

1. For a long time, the difficulties of solving two-dimensional and axially symmetric problems in corpuscular optics were regarded as insurmountable when the space charge was taken into account [1]. The first results were obtained on the basis of the method of similarity and from analogs [1 - 6]. These results also revealed two basic types of problems involving curvilinear beams and provided useful initial information for further mathematical study of these problems. At the present time, the combination of analogs [1 - 6] with electronic computers represents a most essential problem.

2. The studies of one-dimensional corpuscular streams (beams) have been exhausted by the results obtained in [7 - 11]. The question of the necessity of a rigorous formulation of the problem for an electron beam "of more than one dimension" is raised in [12], but no derivation of the systems of equations is given and no boundary conditions are formulated. The absence of a mathematical formulation of this problem can be expressed by the fact that the Langmuir - Child law found fifty years ago [7, 8] represents the entire quantitative basis of modern corpuscular optics in which the space charge is taken into account.

Two-dimensional, axially symmetric, and three-dimensional corpuscular beams with a space charge are characterized by curvilinear trajectories of the particles. They have not yet been studied mathematically. Attempts to apply to them Pierce's scheme [10], which distinguished a region G_0 in the beam (Figure 1) where $\Delta U(x, y) = 4\pi q(U; x, y) \neq 0$ and a region G_1 outside the beam where $\Delta \Psi(x, y) = 0$, have not given any positive results. Pierce's concept of a "laterally bounded beam," which is correct for his problem of a beam of rectilinearly moving particles [10], has no clear meaning, either physical or mathematical. This statement is fully confirmed by experiments. For corpuscular beams with more than one dimension, we eliminate the concept of lateral boundaries. They can be taken conventionally as the "outermost" trajectories confining a given part of the beam only after the distribution of the functions describing the beam and the field (U is the potential, ρ is the space charge density, v_x and v_y are the components of the velocity vector) have been examined and these functions found by a solution of the mathematical problem.

Thus, we shall distinguish:

a) a rectilinear beam having rectilinear trajectories and lateral boundaries formed by the outermost trajectories; this beam is always one dimensional and exists only when special types of fields exist outside it [10]; the concept of a beam is applicable to this type of flow (Figure 1);

b) a curvilinear flow, representing a generalized concept of a beam having curvilinear trajectories of the particle; it has no lateral walls which are defined in a single valued manner; it is characterized only by the distribution of the function U , ρ , v_x , v_y ; it always has more than one dimension.

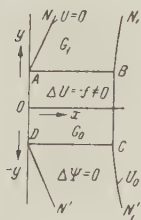


Fig. 1. Rectilinear flow of particles having rectilinear (plane) boundaries.

1. SYSTEM OF DIFFERENTIAL EQUATIONS

3. In the general case, the motion of charged particles in an electromagnetic field can be described by Maxwell's system of differential equations combined with the differential equation of motion of a particle [13]:

$$\text{rot } \vec{E} = \frac{1}{c} \frac{\partial \vec{H}}{\partial t}, \quad (1)$$

$$\text{div } \vec{E} = 4\pi\rho, \quad (2)$$

$$\text{rot } \vec{H} = \frac{1}{c} \frac{\partial \vec{E}}{\partial t} + \frac{4\pi}{c} \vec{I}, \quad (3)$$

$$\text{div } \vec{H} = 0, \quad (4)$$

$$\frac{d\vec{P}}{dt} = e\vec{E} + \frac{e}{c} [\vec{v}_0 \vec{H}], \quad (5)$$

where \vec{E} is the vector of the electric intensity of the field; \vec{H} is the vector of the magnetic intensity of the field; $\vec{I} = \rho \vec{v}$ is the current density vector; \vec{P} , $m\vec{v}_0 / \sqrt{1 - (\vec{v}_0/c)^2}$ is the

momentum vector; and c is the velocity of light in free space. Furthermore, two velocity vectors are distinguished: \vec{v}_0 the velocity of the particle under consideration, and \vec{v} is the velocity of all the remaining particles which are moved by the field. Since the particle velocities in the neighborhood of a given point can be regarded as identical, we shall in the future omit the subscript from \vec{v} in the equation of motion.

For the stationary case, in which there is no external magnetic field, we note that

$$\frac{\partial \vec{H}}{\partial t} = \frac{\partial \vec{E}}{\partial t} = 0, \quad \text{div rot } \vec{H} = \text{div} \left(\frac{4\pi}{c} \vec{I} \right) = 0, \quad (6)$$

Thus, we have a system of differential equations describing a flow of charged particles in an electrostatic field

$$\text{div } \vec{E} = 4\pi\rho, \quad \text{div} (\rho \vec{v}) = 0, \quad \frac{d\vec{P}}{dt} = e\vec{E} + \frac{e}{c} [\vec{v} \vec{H}].$$

For a nonrelativistic stationary flow, if we neglect the magnetic field of the flow of particles, we note that

$$\vec{P} = m\vec{v}, \quad \frac{d\vec{P}}{dt} = m (\vec{v} \nabla) \vec{v}, \quad E = -\text{grad } U, \quad (7)$$

If we transform to the Cartesian coordinate system and use dimensionless variables, we have five equations

$$\frac{\partial^2 \Phi}{\partial \xi^2} + \frac{\partial^2 \Phi}{\partial \eta^2} + \frac{\partial^2 \Phi}{\partial \zeta^2} = \rho_*, \quad (8)$$

$$\frac{\partial \rho_*}{\partial \xi} v_\xi + \frac{\partial \rho_*}{\partial \eta} v_\eta + \frac{\partial \rho_*}{\partial \zeta} v_\zeta = 0, \quad (9)$$

$$v_\xi \frac{\partial v_\xi}{\partial \xi} + v_\eta \frac{\partial v_\xi}{\partial \eta} + v_\zeta \frac{\partial v_\xi}{\partial \zeta} = C_1 \frac{\partial \Phi}{\partial \xi}, \quad (10)$$

$$v_\xi \frac{\partial v_\eta}{\partial \xi} + v_\eta \frac{\partial v_\eta}{\partial \eta} + v_\zeta \frac{\partial v_\eta}{\partial \zeta} = C_1 \frac{\partial \Phi}{\partial \eta}, \quad (11)$$

$$v_\xi \frac{\partial v_\zeta}{\partial \xi} + v_\eta \frac{\partial v_\zeta}{\partial \eta} + v_\zeta \frac{\partial v_\zeta}{\partial \zeta} = C_1 \frac{\partial \Phi}{\partial \zeta} \quad (12)$$

with five unknowns, Φ , ρ_* , v_ξ , v_η , v_ζ , where

$$\Phi = \frac{U}{U_0}; \quad \rho_* = c\rho; \quad v_\xi = \frac{v_x}{v_0}; \quad v_\eta = \frac{v_y}{v_0}; \quad v_\zeta = \frac{v_z}{v_0};$$

$$\xi = \frac{x}{L}; \quad \eta = \frac{y}{L}; \quad \zeta = \frac{z}{L}; \quad U_0 = \text{const}; \quad L = \text{const};$$

$$v_0 = \sqrt{\frac{2e}{m} U_0} = \text{const}; \quad c = -4\pi L^2 / U_0 = \text{const};$$

$$C_1 = \frac{e}{m} \frac{U_0}{v_0^2} = \frac{1}{2}.$$

For an axially symmetric field, the most important case in corpuscular optics, we obtain similarly for equations

$$\frac{\partial^2 \Phi}{\partial R^2} + \frac{1}{R} \frac{\partial \Phi}{\partial R} + \frac{\partial^2 \Phi}{\partial \xi^2} = \rho_*, \quad (13)$$

$$\frac{1}{R} \frac{\partial}{\partial R} (\rho_* R v_R) + \frac{\partial}{\partial \xi} (\rho_* v_\xi) = 0, \quad (14)$$

$$v_R \frac{\partial v_\xi}{\partial R} + v_\xi \frac{\partial v_\xi}{\partial \xi} = C_1 \frac{\partial \Phi}{\partial \xi}, \quad (15)$$

$$v_R \frac{\partial v_R}{\partial R} + v_\xi \frac{\partial v_R}{\partial \xi} = C_1 \frac{\partial \Phi}{\partial R} \quad (16)$$

with four unknowns, Φ , ρ_* , v_ξ , v_R , where

$$\Phi = \frac{U}{U_0}; \quad \rho_* = c\rho; \quad v_\xi = \frac{v_x}{v_0}; \quad v_R = \frac{v_r}{v_0}; \quad R = \frac{r}{L}; \quad \xi = \frac{x}{L}.$$

Here r and x are the dimensionless coordinates of the cylindrical system.

In connection with the elimination of the concept of the boundary of a curvilinear flow (see above), a knowledge of the velocities determining a trajectory loses its basic point, since a definition of a trajectory is not essential here. It is necessary to know only the potential distribution Φ and the components of the current density $I_x = \rho v_x$, $I_y = \rho v_y$, $I_z = \rho v_z$.

2. BOUNDARY CONDITIONS AND TYPES OF PROBLEMS

4. To formulate the boundary conditions, specific problems must be considered. For convenience in our treatment, we shall distinguish two basic types of problems depending on whether the following characteristic conditions are contained in a set of boundary conditions.*

*The terms "direct" problem and "converse" problem were used in the works of G. A. Grinberg, but in a diametrically opposed sense, and only for special problems in corpuscular optics in which space charge was ignored.

I. The direct problem. A system of external electrodes is given, but the distribution of the function is not given in any single section of the flow.

II. The converse problem. The system of external electrodes is not given (see Section 7), but the distribution of the functions is given in one section of the flow, for instance, at the collector.

In both cases, the unknowns which must be determined are the distributions of the functions φ , ρ , v_x , and v_y . For simplicity, we shall confine ourselves to a study of optical systems with two electrodes defined by the boundary conditions.

5. We are considering problems with a stationary nonrelativistic flow of electrons or ions (in a vacuum) under the action of an electrostatic field formed by a potential difference applied between two electrodes NN' and $N_1N'_1$ (Figs. 2 and 3), and the field of the space charge of the beam. Only the part AD of the electrode NN' emits particles. The flow of particles moving from NN' to $N_1N'_1$ forms an electric current $\vec{I} = \rho \vec{v}$.

We shall assume that all the charges in the flow of particles are of the same sign, so that the total charge is not equal to zero. We shall also assume that all the charges are identical. We shall take the initial velocities of the particle equal to zero. Accordingly,

$$\frac{mv^2}{2} = eU. \quad (17)$$

We shall suppose that the emissivity of the portion AD of the electrode NN' is not restricted. For this, we have the condition

$$(\rho_* \varphi^{1/2})_{\varphi \rightarrow 0} \rightarrow \text{const} \begin{cases} \neq 0, \\ \neq \infty. \end{cases} \quad (18)$$

Let us consider one typical example of a problem and its converse.

3. DIRECT PROBLEM

6. On the plane (ξ, η) , there is an open region G contained between two infinite parallel straight lines $NN' \parallel N_1N'_1$ (Figure 2), separated by a distance $OE = 1$.

In the region G (ξ, η) , the field of the function ρ_* , φ , v_x , v_y , which is described by a system of differential equations of the form (8) - (12) - for the case of a two-dimensional field, is distributed symmetrically about the ξ axis (we take the origin of the Cartesian coordinates at the point O).

The following boundary conditions are given:

- 1) on the straight line NN' , $\varphi(0, \eta) = 0$
- 2) on the straight line $N_1N'_1$, $\varphi(1, \eta) = 1$
- 3) on the segment AOD (which is symmetric with respect to the point O)

$$(\varphi^{1/2} \rho_*)_{\varphi \rightarrow 0} \rightarrow \text{const} \begin{cases} \neq 0, & \frac{\partial \varphi}{\partial \xi} \\ v_x(\xi, \eta) \rightarrow 0, \\ \neq \infty, & v_y(\xi, \eta) \Big|_{\xi \rightarrow 0} \end{cases}$$

4) the conditions at infinity

$$\frac{\partial \varphi}{\partial \xi} \Big|_{|\eta| \rightarrow \infty} \rightarrow 1; \quad \frac{\partial \varphi}{\partial \eta} \Big|_{|\eta| \rightarrow \infty} \rightarrow 0.$$

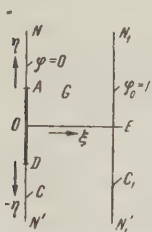


Fig. 2. The direct problem. A curvilinear two-dimensional flow.

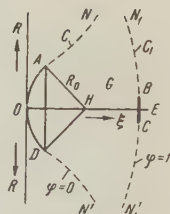


Fig. 3. The converse problem. A curvilinear axially symmetric flow.

It is required to find the distribution of the function $\varphi(\xi, \eta)$, $\rho_*(\xi, \eta)$, $v_z(\xi, \eta)$, $v_n(\xi, \eta)$ in the region G.

4. CONVERSE PROBLEM

7. In the space (ξ, R) , there is an open region G, which is symmetric with respect to the ξ -axis and which is contained between two axially symmetric and bounded surfaces C and C_1 , separated by a distance $OE = 1$ at their centers (see Figure 3).

The central part of C is the surface of the segment AOD of a sphere with a center H on the ξ -axis and with a radius $R_0 = AH = \text{const}$. The radius of the segment is $r_0 = AD/2 = \text{const}$. The central part of the boundary surface is a circle of the radius $r_0 = AD/2 = \text{const}$. The outer parts of the boundary surfaces C and C_1 are unknown.

The field of the function φ , ρ_* , v_z , and v_R , which is described by the system of differential equations (13) - (16), is distributed in the region G. The following boundary conditions are given:

- 1) on the surface C, $\varphi(\xi, R) = 0$,
- 2) on the surface C_1 , $\varphi(\xi, R) = 1$,
- 3) when the spherical surface AOD is approached,

$$(\varphi^{1/2} \rho_*)_{\varphi \rightarrow 0} \rightarrow \text{const} \left\{ \begin{array}{l} \neq 0, \quad \frac{\partial \varphi}{\partial n} \\ v_R(\xi, R) \\ \neq \infty, v_z(\xi, R) \end{array} \right|_{\varphi \rightarrow 0} \rightarrow 0,$$

- 4) the distribution of the functions in the circle BEC (for $0 \leq R \leq r_{01}$)

$$\rho_* \equiv \rho_*(1, R), \quad v_z \equiv v_z(1, R), \quad v_R \equiv v_R(1, R).$$

- 5) the conditions at infinity

$$\rho_*(\xi, R) \left| \begin{array}{l} \frac{\partial \varphi}{\partial R} \end{array} \right|_{R \rightarrow \infty} \rightarrow 0.$$

It is required to find the distribution of the functions $\varphi(\xi, R)$, $\rho_*(\xi, R)$, $v_z(\xi, R)$, $v_R(\xi, R)$ in the region G and thus to find the system of external electrodes: the electrodes in the region G, which correspond to the extremum of the potential (including $\varphi = 0$, $\varphi = 1$), are necessary for the knowledge of the distribution $\varphi(\xi, R)$ which we have found and of the other functions.

The problems which we have formulated represent new types of boundary problems, which have not yet been studied by mathematicians.* It is important that these problems attract the attention of mathematicians and theoretical physicists.

We express our deep gratitude to A.A. Dorodnitsyn for his interest in the work and for a number of valuable comments. We are also sincerely grateful to N.N. Moiseyev for his interesting discussions and comments.

REFERENCES

1. V.M. Breytman, Methods of computing ionic optical systems with the space charge taken into account, Doklady AN SSSR, 1959, vol. 127, 6, 1187.
2. V.M. Breytman, Mathematical basis of an analog method for a two-dimensional field described by the generalized Poisson equation on the surface of a three-dimensional electrolyte, Izv. Vuzov MVO SSSR (Energetika), 1960, 3, 9, 46.
3. V.M. Breytman, Examples of practical computations of electronic (ionic) optical systems in which the space charge is taken into account by an analog on a three-dimensional electrolyte, Radiotekhnika i Elektronika, 1960, 5, 9, 1484.

*They are problems of a mixed type, generalizing the problems of the elliptical and hyperbolic types.

4. V.M. Breytman, Computations of the field of a ribbon-shaped beam which diverges under the action of a space charge in a plane parallel diode with a finite width, *Radio-tehnika i Elektronika*, 1960, 5, 10, 1619.
5. V.M. Breytman, Similarity in physical phenomena from the geometric point of view, *NDVSh (Power)*, 1958, 1, 98.
6. V.M. Breytman, Dimensional and analog methods, *Trudy mezhvuzovskoy konferentsii po vopr. fiz. i matem. modelirovaniya*, Dokl. F-38, MEI 1959.
7. D.C. Child, Discharge from hot CaO, *Phys. Rev.*, 1911, 32, 5
8. I. Langmuir, The effect of space charge and residual gases on thermionic currents in high vacuum, *Phys. Rev.*, 1913, ser. 2, 2.
9. I. Langmuir, K. Blodgett, *Phys. Rev.*, 1923, 22, 347; 1924, 24, 49.
10. J.R. Pierce, Rectilinear electron flow in beams, *J. Appl. Phys.*, 1940, 11, 548.
11. R. Helm, et al, *Electr. commun.*, 1947, 24, 101.
12. K. Spangenberg, *F. Franklin Inst.*, 1941, 232, 365.
13. L.D. Landau and E.M. Lifshits, *Field Theory*, GTTI, 1959.

Received February 27, 1960

A PULSE METHOD OF DETERMINING THE PARAMETERS OF DRIFT TRANSISTORS

V. I. Shveykin

This article gives a basis for the pulse method of measuring the charge proposed in [1]. This method makes it possible to determine the frequency properties of drift transistors, the effective lifetime of minority carriers in the base, and the values of the diffusion capacitance and static capacitance of the emitter and the collector.

INTRODUCTION

The pulse method which is proposed in [1] for measuring the charge of minority carriers accumulated in the base of junction transistors in the process of current transmission from the emitter to the collector is in theory applicable to any transistor, whether of the alloy-type or the drift-type. As distinguished from alloy-type transistors, drift transistors are characterized by an electric field in the base and a small value of the base charge. The first of these properties of drift transistors makes it necessary to establish the relationship between the base charge and the frequency parameters of the transistor when there is an electric field in the base, while the second property makes it necessary to examine the possible errors which can occur in a real measuring circuit (Figure 1) and to take into account the effect of the capacitances of the emitter and collector junctions of a transistor on the measured base charge.

The study of these questions which is conducted in this article permits us to assert that the simple and reliable pulse method of measuring the base charge can be used to determine rapidly and with sufficient accuracy a number of the most important characteristics of both alloy-type and drift transistors.

1. THE RELATIONSHIP BETWEEN THE BASE CHARGE AND THE FREQUENCY PARAMETERS OF DRIFT TRANSISTORS

To characterize the frequency properties of drift transistors, the frequency at which the real part of the amplification coefficient for the current α is 1/2 its low-frequency value α_0 is used [2, 3]. The chief advantage of ω_1 over the critical frequency of the amplification coefficient for the current ω_α consists in the fact that at the frequency ω_1 the absolute value of the amplification coefficient for the current in a circuit with a common emitter is

$$|\beta| \simeq 1 \quad (1)$$

for any distribution of the impurity in the region of the base [3]. The approximate equation (1) becomes an exact equation if we assume $\alpha_0 = 1$. For real transistors, this is accurate to a sufficient degree.

We shall show that for junction transistors with any type of distribution of the impurity in the region of the base this frequency ω_1 is inversely proportional to the mean free time of the carriers t_{pr} , i. e.,

$$\omega_1 = \frac{1}{t_{pr}} = \frac{I_e}{Q} \quad (2)$$

here Q is the steady-state charge of the minority carriers when an emitter current I_e flows [1].

If we assume that the transistor is one-dimensional when operating as an amplifier, we can write

$$\frac{dp(x)}{dx} - \Lambda E(x) p(x) = -\frac{I_p(x)}{qDS}, \quad (3)$$

where $p(x)$ is the concentration of minority carriers (holes) in the base; $\Lambda = q/kT$; D is the diffusion coefficient of the holes in the base; S is the area of the junction; $I_p(x)$ is the hole current at the point x ; and $E(x)$ is the value of the electric field in the base of the triode.

The solution to (3) with the boundary condition $p(x)|_{x=w} = 0$ at the collector junction (correct in the amplification regime) has the following form:

$$p(x) = \frac{1}{qD} e^{-\int_x^w \Lambda E(\xi) d\xi} \left[\int_x^w I_p(\xi) e^{\int_\xi^w \Lambda E(\xi) d\xi} d\xi \right], \quad (4)$$

where w is the thickness at the base. Integrating (4) along the entire base region and assuming $I_p(\xi) = I_e = \text{const}$, i. e., neglecting the phenomenon of recombination, we obtain the following expression for the mean free time of the holes for passage through the base region with any electric field $E(x)$ in the base region:

$$t_{pr} = \frac{Q}{I_e} = \frac{1}{D} \int_0^w e^{-\int_x^w \Lambda E(\xi) d\xi} \left[\int_x^w e^{\int_\xi^w \Lambda E(\xi) d\xi} d\xi \right] dx. \quad (5)$$

If we let

$$\frac{2}{w^2} \left\{ \int_0^w e^{-\int_x^w \Lambda E(\xi) d\xi} \left[\int_x^w e^{\int_\xi^w \Lambda E(\xi) d\xi} d\xi \right] dx \right\} = v, \quad (6)$$

we can rewrite (5) in the form

$$t_{pr} = \frac{w^2}{2D} \frac{1}{v}. \quad (7)$$

For the relationship between t_{pr} and ω_1 , we shall use the results obtained in [4] for α with an arbitrary field distribution in the base region:

$$\alpha \simeq \operatorname{sch} \sqrt{j \frac{\omega w^2}{Dv}}. \quad (8)$$

We rewrite (8) in the form

$$\alpha = \frac{1}{\operatorname{ch} a \cos a + j \operatorname{sh} a \sin a}, \quad (9)$$

where $a = \sqrt{\omega w^2 / 2Dv}$, and divide the real and imaginary parts of (9). Then from the condition $\operatorname{Re}[\alpha(j\omega)] = 1/2$ we obtain

$$\omega_1 = \frac{1.997 D}{w^2} v \simeq \frac{2D}{w^2} v. \quad (10)$$

If we compare this equation for ω_1 with (7), we will obtain a proof of the initial equation (a). Thus, the charge of the minority carriers in the base of a drift transistor determines the frequency ω_1 in a single-valued manner or

$$f_1 = \frac{I_e}{2\pi Q} = \frac{1}{2\pi t_{pr}} \quad (11)$$

independently of the type of distribution of the impurity in the base region.

The quantity v which appears in the equation for t_{pr} and ω_1 is a very important parameter of drift transistors. In particular, as seen from (7) and (10), it is the factor showing the decrease of t_{pr} , and, correspondingly, the increase of the frequency f_1 for a drift transistor with a field in its base relative to a transistor in which all the conditions are identical, except that there is no field. We shall call v the drift factor.

If an electron current $I_n(x) = 0$ is assumed along the entire base region, i.e.,

$$\frac{dn(x)}{dx} + \Lambda E(x)n(x) \equiv 0,$$

where $n(x)$ is the concentration of electrons in the base, and if we set $n(x) \simeq N_d(x)$, where $N_d(x)$ is the concentration of impurity atoms in the base region, we are able to write the following equation for v [see (6)]:

$$v = \frac{2}{w^2} \int_0^w \frac{1}{N_d(x)} \left[\int_x^w N_d(\xi) d\xi \right] dx. \quad (12)$$

It follows from (12) that, in the absence of drift motion of carriers in the base [$N_d(x) = \text{const}$] $v^{(0)} = 1$. In this case, a measurement of the base charge Q makes it possible to determine both the frequency

$$f_1^{(0)} = \frac{D}{\pi w^2} = \frac{I_e}{2\pi Q}, \quad (13)$$

and the critical frequency

$$f_\alpha^{(0)} = \frac{1.215 D}{\pi w^2} = \frac{1.21 I_e}{2\pi Q}. \quad (14)$$

A comparison of (13) and (14) shows that in an alloy-type transistor there is a simple relationship between f_1 and f_α

$$K = \frac{f_\alpha^{(0)}}{f_1^{(0)}} = 1.21.$$

In the general case, the proportionality factor between f_α and f_1 depends on the distribution function of the impurity in the base. Let us imagine a transistor model with a constant field E_0 in the base region, in which the impurity concentration in the base changes exponentially

$N_d(x) = N_{d0} e^{-\Lambda E_0 x} = N_{d0} e^{-\frac{\Delta V}{kT} \frac{x}{w}}$. Then, it follows from (12) that the drift factor is

$$v(E_0) = \frac{\left(\frac{\Delta V}{kT}\right)^2}{2 \left[\frac{\Delta V}{kT} - 1 + e^{-\frac{\Delta V}{kT}} \right]},$$

and that, consequently, the frequency $f_1(E_0)$ is determined as follows:

$$f_1(E_0) = \frac{D}{\pi w^2} \frac{\left(\frac{\Delta V}{kT}\right)^2}{2 \left[\frac{\Delta V}{kT} - 1 + e^{-\frac{\Delta V}{kT}} \right]}.$$

This result is also obtained in [2] by direct computation with an error smaller than 1%. In this case the coefficient $K = f_\alpha(E_0)/f_1(E_0)$ for all real values for germanium transistors $0 \leq \Delta V/kT \leq 8$ is a linear function of $\Delta V/kT$ [2]:

$$K = 1.21 + 0.09 \frac{\Delta V}{kT}. \quad (15)$$

It follows from (15) that the coefficient K does not depend strongly on $\Delta V/kT$. For the most probable values, $\Delta V/kT \approx 4-5$. Then, $f_\alpha \approx 1.6f_1$. The use of this relation to compute f_α from the measured value of $f_1 = I_c/2\pi Q$ for any value $0 \leq \Delta V/kT \leq 8$ will lead to an error not exceeding 25%.

2. THE POSSIBILITY OF MEASURING SMALL CHARGES

The value of the base charge in drift transistors is approximately two orders of magnitude smaller than an alloy-type transistor, and amounts to 10^{-11} to 10^{-10} k for $I_e = 5-20$ ma.

When the charge is measured in the circuit, shown in Figure 1, the chief errors will arise as a result of the following phenomena: the finite switching time of the measuring diodes D_2 and D_1 (t_{switch}); the finite duration of the trailing edge of the emitter (t_{eff}); and the existence of stray circuit capacitances (C_1, C_2, C_3).

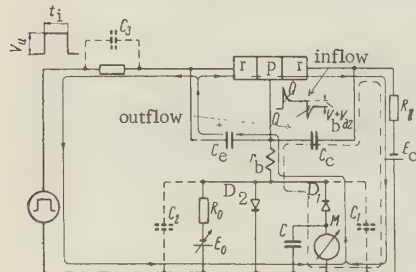


Fig. 1. Measuring circuit for charge in which the capacitances of the emitter and the collector of the triode, as well as the stray capacitances of the circuit are taken into account.

(the quality factor of the diode [5]). It characterizes single-valuedly its frequency properties [6] (compare with t_{pr} for triodes). The condition which must be satisfied for the finite duration t_{switch} to be neglected is $Q_{\text{outflow}} = Q = I_e t_{pr} \gg Q_{D2} = I_{b0} t_0$, or

Let us consider the effect of t_{switch} . If the transistor is ideal, having a base current $I_{b0} = 0$ and a fully compensated current I_{c0} [1], it is obvious that we can assume with a sufficient degree of accuracy that Q_{inflow} passes entirely through the diode D_2 ; while Q_{outflow} passes entirely through the diode D_1 and the instrument M (see Figure 1). Of course, $I_c \gg I_{SD2}, I_{SD1}$, where I_{SD1} and I_{SD2} are the back currents of the diodes D_1 and D_2 .

In real transistors, $I_{b0} \neq 0$, and at the time when the current pulse I_e of the emitter ceases ($t = t_i$) a current I_{b0} passes through the diode D_2 . In this case, the charge registered by the instrument will certainly not be smaller than Q_{outflow} minus Q_{D2} , where $Q_{D2} = I_{b0} t_0$ is the charge stored in the diode D_2 by the time t_i . The quantity t_0 is the time constant of the diode

$$t_0 \ll \beta_0 t_{pr}, \quad (16)$$

where β_0 is the low-frequency value of β .

Since $\beta_0 \gg 1$ for real curves, it follows from (16) that the effect of the finite t_{switch} can be made negligible by selecting diodes with frequency properties as good as those of the transistors which are being measured.*

A condition analogous to (16) can be obtained for the duration t_{eff} (we give it here without proof):

$$t_{eff} \ll \beta_0 t_{pr} = \tau, \quad (17)$$

where τ is the effective lifetime of the holes in the base. For drift transistors, the values of τ usually lie within the limits of 0.2–1 microseconds. Therefore, the error resulting from the finite duration t_{eff} can be kept below 5–10% if t_{eff} is equal to one hundredths or a few hundredths of a microsecond, as is quite feasible in practice.

It follows from the above discussion that the capacitance $C_{1,2} = C_1 + C_2$ (basically this is the capacitance of the diode D_1 and D_2) can decrease the charge $Q_{outflow} = Q$ by an amount of only $Q_{C_{1,2}} = C_{1,2} V_{D_2}$, where V_{D_2} is the voltage across the diode D_2 at the time $t = t_i$.

If we suppose that $C_{1,2} = 1 \mu\text{f}$ for point-contact diodes D_1 and D_2 , and that $V_{D_2} = 0.3\text{v}$, we obtain $Q_{C_{1,2}} = 0.3 \cdot 10^{-2} k$. This is considerably lower than the measured charges.

The capacitance C_3 will have comparatively no effect if measures to prevent the breakdown of the emitter junction are taken.

Summing up our discussion, we conclude that, if the diodes D_1 and D_2 are suitably chosen, (16) if the trailing edge of the rectangular-pulse generator is sufficiently steep (17), and if the measuring circuit is carefully assembled, this method will make it possible to measure charges of the order of $10^{-11} k$ and above with an error not exceeding 10–15%.

3. THE EFFECT OF CAPACITANCES OF THE EMITTER AND COLLECTOR JUNCTION ON THE MEASURED VALUE OF THE BASE CHARGE. EXPERIMENTAL RESULTS OF THE MEASUREMENT OF DRIFT TRIODE PARAMETERS BY THE PULSE METHOD

For alloy-type transistors in which, under typical operation, the charge accumulated in the base was large relative to the charges stored in the capacitances of the emitter C_e and the collector C_c , it was permissible to neglect the effect of these capacitances in [1]. However, when the charge is measured in the base of drift transistors, this factor cannot be ignored, especially since these charges may have a determining effect at small currents I_e .

Figure 1 shows the circuit for measuring the charge when the capacitances C_e and C_c of the measured transistor are taken into account. To be specific, we shall assume that the transistor is operating in the active region ($I_e \leq E_c / \alpha_0 R_l$) and we shall suppose that the duration of the current pulse of the emitter t_i is sufficient for all the transient processes associated both with the accumulation of charge into the base and with the accumulation of charge in the capacitances C_e and C_c to settle. In this case, a charge

$$Q_e = \bar{C}_e V_e,$$

is accumulated in the capacitance C_e by the time t_i . Here, $\bar{C}_e = \frac{1}{V_e} \int_0^{V_e} C_e(V) dV$ and V_e is the voltage at the emitter junction when a current I_e passes.

From the circuit of Figure 1 it is easily seen that the discharge circuit of the capacitance C_c , indicated in Figure 1 by a solid line, is short-circuited through the measuring instrument, increasing its readings of the charge [1] by Q_e .

*For the diodes D_1 and D_2 , we used D10 B diodes. It is possible to use crystal detectors operating in the range of centimeter wavelengths. An "ideal measuring diode can be provided by a tunnel diode with a degenerate characteristic in the reverse direction only (a reverse tunnel diode).

The total change of the charge in the capacitance $C_c = C_{sc} + C_{dc}$, where C_{sc} is the static capacitance and C_{dc} is the diffusion capacitance of the collector, during the transient period of the emitter, current is equal to

$$Q_R = \int_{V_R}^{E_C} [C_{sc}(V) + C_{dc}(V)] dV. \quad (18)$$

Here, $V_c = E_c - I_e R_l \alpha_0 + V_b + V_{D2}$; E_c is the voltage of the collector battery, R_l is the load resistance; V_b is the voltage produced by the base current I_{b0} across the base resistance r_b ; and V_{D2} is the voltage across the diode D_2 (see Figure 1). This is the charge which will be registered on the measuring instrument. The path of the discharge of the capacitance C_c in Figure 1 is indicated by the dot- and dash-line. For the case where the collector is short-circuited ($R_l = 0$), the charge stored in the capacitance C_c is equal to

$$Q'_c = \int_{E_c + V_b + V_{D2}}^{E_C} C_c(V) dV \simeq C_c(E_c) [V_b + V_{D2}]$$

and it is easily seen that the readings of the instrument will be decreased by this same amount. Thus, when the collector short-circuits, as must be the case in the determination of the critical frequency, the readings of the instrument, with the effect of C_e and C_c taken into account, will be

$$Q_{pr} = Q + Q_e - Q'_c$$

or

$$Q_{pr} = I_e \left\{ t_{pr} + \frac{\bar{C}_e}{\Delta I_e} \ln \left(\frac{I_e}{I_{se}} + 1 \right) - C_c \left[\frac{r_b}{\beta_0} + \frac{1}{\Delta I_e} \ln \left(\frac{I_e}{\beta_0 I_{SD2}} + 1 \right) \right] \right\}, \quad (19)$$

where I_{se} and I_{SD2} are the back currents of the emitter junction and diode D_2 , respectively.

Setting $\bar{C}_e = 50 \mu\text{mf}$, $I_{se} = 10^{-6}$ amp, and $t_{pr} = 2.5 \cdot 10^{-9}$ sec (this corresponds to the value $f_1 \simeq 65 \text{ Mc}$), we obtain for the condition $Q = Q_e$ the value of I_e^* at which the charges Q and Q_e are commensurable. For domestic transistors of the P401 - P403 type, this value of I_e^* is equal to 4 ma for selected \bar{C}_e , t_{pr} , and I_{se} .

Let us now determine the contribution of the capacitance C_c relative to the capacitance C_e for currents $I_c \leq I_e^* = 4$ ma. Considering that $I_{se} = I_{SD2} = 10^{-6}$ amp and $r_b = 100$ ohms, and setting $\beta = 50$, we obtain

$$\frac{Q_e}{Q_c} \simeq 1.8 \frac{\bar{C}_e}{C_c(E_e)}.$$

Since the value of $C_c(E_c)$ is approximately one order of magnitude smaller than that of \bar{C}_e in drift transistors, the contribution of the capacitance C_c will be 5 - 10% of the contribution of the capacitance C_e , for currents $I_e < 4$ ma.

For the above discussion, it follows, that if the experimental value of Q_{pr} depends on $\ln I_e$ linearly for currents $I_e < 4$ ma, the slope of this straight line with respect to $\ln I_e$ will determine the value of the capacitance \bar{C}_e , which can be regarded as constant in this interval. Figure 2 contains experimental curves for the dependence of Q_{pr} on $\ln I_e$ for transistors of the P401, P402

and P403 types. Indeed, it can be seen from Figure 2 that for emitter currents smaller than 2 - 3 ma the dependence of Q_{pr} on $\ln I_e$ is nearly linear (the charge Q_e increases logarithmically with the current I_e . At large values of I_e , the charge Q_{pr} increases rapidly as a

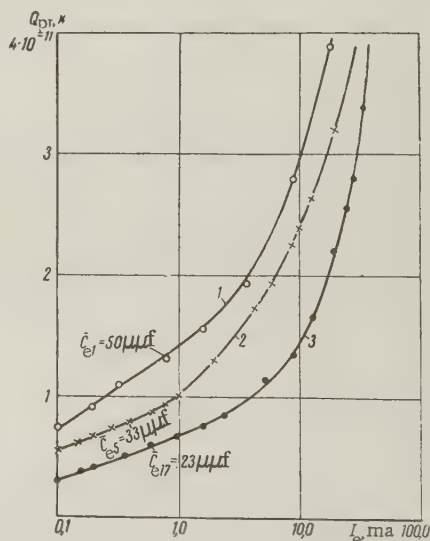


Fig. 2. Experimental curve for Q_{pr} as a function of $\ln I_e$. Measuring condition: $E_c = 5$ v, $t = 1 \mu$, and $T = 2 \mu$.

1) P402 No. 1; 2) P403 No. 5; 3) P402 No. 17.

result of the base charge Q which increases linearly with the current I_e . The results of the measurements make it possible to determine \bar{C}_e for a given value of $V_e(I_e)$. The values of \bar{C}_e obtained from the tested transistors are shown in Figure 2. The measurements of $V_e(I_e)$ were made with direct current.

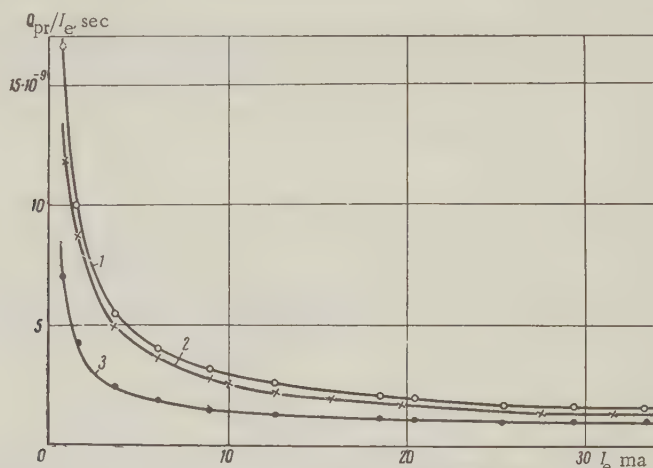


Fig. 3. Experimental curves showing the dependence of Q_{pr}/I_e on I_e . Measuring conditions: $E = 5$ v, $t = 0.5$ μ sec, $T = 1$ μ sec.

1) P402 No. 1; $t_{pr1} = 1.56 \cdot 10^{-9}$ sec ($f_1 = 100$ Mc); 2) P403 No. 5, $t_{pr5} = 1.27 \cdot 10^{-9}$ sec ($f_1 = 125$ Mc); 3) P401 No. 17, $t_{pr17} = 1.0 \cdot 10^{-9}$ sec ($f_1 = 160$ Mc).

For large emitter currents ($I_e \gg I_e^*$), it can be seen from (19) that $r_e = \frac{I}{\Delta I_e} \ln \left(\frac{I_e}{I_{se}} + 1 \right) \rightarrow$

and $r_{D2} = \frac{1}{\Delta I_e} \ln \left(\frac{I_e}{\beta_0 I_{SD2}} + 1 \right) \rightarrow 0$ and that Eq. (19) becomes

$$Q_{pr} = I_e \left(t_{pr} - \frac{r_b C_c}{\beta_0} \right). \quad (20)$$

The effect of $r_b C_c / \beta_0$ can be determined on the basis of the following considerations. For P401 - P403 transistors, the product $r_b C_c$ is guaranteed not to exceed $3.5 \cdot 10^{-9}$ sec. If it is assumed that $\beta_0 = 50$, then $r_b C_c / \beta_0 = 7 \cdot 10^{-11}$ sec, and the correction to the quantity $t_{pr} = 2.5 \cdot 10^{-9}$ sec ($f_1 = 65$ Mc) will amount to less than 3%.

Thus, if the charge is measured with a sufficiently large emitter current I_e , it will be possible to determine the mean free time of passage of the minority carriers through the base region, and, consequently, the frequency f_1 which is related single-valuedly to the mean free time by eq. (11).

Figure 3 contains an experimental curve showing Q_{pr}/I_e as a function of the emitter current I_e . The results show that, as the current I_e increases up to 5 - 10 ma, the ratio Q_{pr}/I_e , which has as its limiting value the mean free time of the minority carrier t_{pr} .

The measured values of t_{pr} , and consequently of f_1 , are shown in Figure 3. The frequencies f_1 are the results of only the internal mechanism of current transmission in drift transistors and are free from the effect of the emitter and collector capacitances.

It should be noted that a knowledge of t_{pr} makes it possible to compute the diffusion capacitance of the emitter junction

$$C_{de} = \frac{dQ}{dV_e} = \frac{dQ}{dI_e} \frac{dI_e}{dV_e} \simeq t_{pr} I_e \Delta = Q \Delta.$$

Thus, for instance, for the P401 No. 17 triode (see Figure 3), $t_{pr} = 1$ μ sec. For a current $I_e = 1$ ma, the diffusion capacitance of the emitter is equal to $C_{de} = 40$ μ mf.

It is also possible to determine the value of the diffusion capacitance of the collector $C_{dc} = dQ/dE_c \simeq \Delta Q / \Delta E_c$ by measuring the decrease of the base charge ΔQ which takes

place when the collector voltage increases by ΔE_C (other conditions remaining the same) as a result of modulation of the width of the base region (the resistance R_l is equal to zero in this measurement).

For instance, for the P402 No. 1 transistor, the value of C_{dc} is equal to $0.18 \mu f$ at a current $I_e = 20$ ma and with $\Delta E_C = 10$ v - 5 v = 5 v.

In the case where a load resistance R_l is connected in the collector circuit, the effect of the capacitance C_c can be very large, since the charge Q_c [see (18)] may be comparable with or even larger than charge Q at sufficiently large values of $I_e R_l \alpha_0$. When there is a load R_l ,

the readings of the instrument increase by an amount $Q_c = \int_{E_C - I_e \alpha_0 R_l}^{E_C} C_c(V) dV$ relative to the

case where there is a short-circuit in the collector circuit. The condition $E_C \gg V_b + V_{D2}$ is assumed to be satisfied. The difference between the measured values for the two cases $R_l \neq 0$ and $R_l = 0$, other conditions being equal, gives the quantity $Q_c = Q_{pr}/R_l \neq 0$. Consequently, it makes it possible to determine the average value of the sum of the static capacitance and the diffusion capacitance of the collector in the range of voltages from $E_C - I_e \alpha_0 R_l$ to E_C .

$$\bar{C}_c = \frac{Q_c}{I_e R_l \alpha_0}.$$

If the voltage of the collector battery is adjusted at $E_C - I_e R_l \alpha_0$ simultaneously with the short-circuiting of the load R_l , the difference between the charges measured in this case $Q_{pr} \Big|_{R_l \neq 0}^{E_C} - Q_{pr} \Big|_{R_l = 0}^{E_C - I_e R_l \alpha_0}$ will be equal to the value of the charge stored in the static

capacitance of the collector junction alone. Consequently, for measured $Q_{pr} \Big|_{R_l = 0}^{E_C} - Q_{pr} \Big|_{R_l = 0}^{E_C - I_e R_l \alpha_0}$, $I_e R_l \alpha_0$ the average value of the static capacitance of the

collector junction C_{sc} in the range of voltages from $E_C - I_e R_l \alpha_0$ to E_C can be computed from the equation

$$\bar{C}_{sc} = \frac{1}{I_e R_l \alpha_0} \int_{E_C - I_e R_l \alpha_0}^{E_C} C_{sc}(V) dV = \frac{Q_{pr} \Big|_{R_l \neq 0}^{E_C} - Q_{pr} \Big|_{R_l = 0}^{E_C - I_e R_l \alpha_0}}{I_e R_l \alpha_0}. \quad (21)$$

For drift transistors, the dependence of C_{sc} on the applied voltage V can be well described by the equation [7]

$$C_{sc}(V) = AV^{-1/2}. \quad (22)$$

A substitution of (22) into (21) gives

$$\bar{C}_{sc} = C_{sc}(E_C) \left[1 + \frac{1}{6} \frac{I_e R_l \alpha_0}{E_C} \right], \quad (23)$$

which is obtained by replacing the binomial $(E_C - I_e R_l \alpha_0)^{3/2}$ with the first three terms of its expansion. It is obviously assumed that $E_C > I_e R_l \alpha_0$.

In the case where $E_C = I_e R_l \alpha_0$, a direct computation of (21) leads to the following simple \bar{C}_{sc} and $C_{sc}(E_C)$:

$$\bar{C}_{sc} = 1.5 C_{sc}(E_C).$$

For sharp (alloy-type junctions, for which the equation $C_c(v) = AV^{-1/2}$ holds, the corresponding equation has the form

$$C_{sc} = 2C_{sc}(E_C).$$

Figure 4 shows the dependence of $[1/C_{SC}(E_C)]^3$ on E_C for transistors of the P401 and P402 types. The values of $C_{SC}(E_C)$ were determined from (23) after C_{SC} had been measured by the method previously indicated. The results obtained (see Figure 4) indicate that eq. (22) holds. This is in agreement with [7].

When the base charge of a saturated transistor Q_{sat} is measured by the method described in [1], it is necessary to take into account the charge stored in C_{SC} and C_E . A knowledge of Q_{sat} makes it possible to determine the average lifetime of the holes when the base changes from zero to I_b :

$$\bar{\tau} = \frac{Q_{sat}}{I_b} = \frac{1}{I_b} \int_0^{I_b} \tau(I_b) dI_b.$$

The junction capacitances of the transistor can be ignored if $\tau(I_b)$ is determined from the equation

$$\tau(I_b) = \frac{Q_2 - Q_1}{I_{b2} - I_{b1}} = \frac{\Delta Q_{sat}}{\Delta I_b},$$

where Q_2 and Q_1 are the instrument readings corresponding to two nearby values of the base current I_{b1} and I_{b2} for a saturated transistor. This is the result of the fact that, when the base current changes from I_{b1} to I_{b2} , the charge in C_{SC} and C_E changes quite negligibly relative to the change of Q_{sat} of the transistor. The change in the base current $\Delta I_b = I_{b2} - I_{b1}$ can be made either by changing the emitter current by ΔI_e , or by changing the collector current ΔI_c through a change in R_L . In the first case, $\Delta I_b \simeq \Delta I_e$, while in the second case $\Delta I_b \simeq \Delta I_c$. Figure 5 shows values of $\tau(I_b)$ measured in this way at the values $I_b = 10$ ma. The results indicate that the effective lifetime of the holes in the base of the saturated drift transistor increases with increasing base current.

This pulse method of determining the parameters of transistors by measuring the charge is simple and reliable and has the following advantages over ordinary measuring methods:

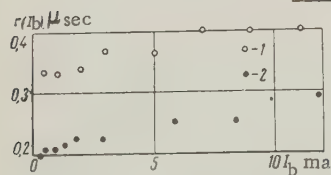


Fig. 5. Experimental curves showing τ as a function of I_b . Measuring conditions: $E_C = 10$ v, $R_L = 3.3$ kilohms, $t_i = 1 \mu sec$, $T = 2 \mu sec$. 1) P402 No. 1; 2) P401 No. 17.

1) Very frequently transistors are used in pulse operation in which the power dispersed in a pulse exceeds the tolerated limit. Ordinary measurements of the parameters of transistors are made under constant conditions. This does not make it possible to take measurements under conditions exceeding the limit. The method we have proposed is free from this disadvantage: the ratio $\kappa = T/t_i$ of the generator at the input emitter current pulses can always be selected sufficiently large.

2) As the emitter current I_e increases, the base charge Q also increases. Consequently, the smallest current measured according to the circuit of Figure 1 can always be achieved and exceeded by increasing the current I_e . Without there being any danger that the transistor will be damaged through overheating, the necessary increase of the current I_e can always be obtained if the "porosity" of the rectangular pulses is sufficiently large. This obviously makes it possible to extend the limit of measurements of critical frequency

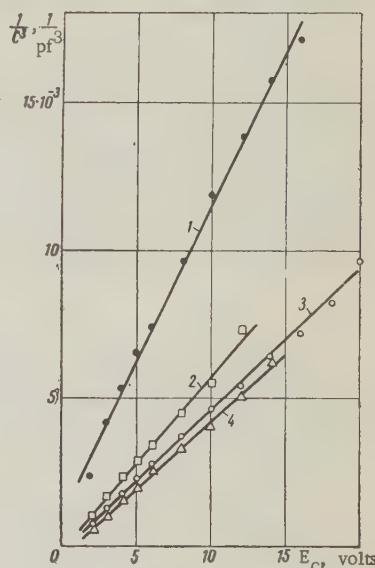


Fig. 4. Experimental curve showing $1/C_{SC}^3$ as a function of E_C . Measuring conditions: $I_e = 0.8$ ma, $R_L = 2.2$ kilohms, $t_i = 0.5 \mu sec$, $T = 1 \mu sec$.

1) P401 No. 17; 2) P402 No. 14; 3) P402 No. 1; 3) P401 No. 21.

considerably beyond that of ordinary methods. Thus, for instance, for the P410 No. 22 transistors, the measured value of the charge at $I_e = 35$ ma was equal to $1.4 \cdot 10^{-11}$ k, which corresponds to $f_1 \approx 400$ Mc. This measurement was made with $t_1 = 0.4$ μ sec, $T = 1$ μ sec, $E = 5$ v.

3) The influence of $r_b C$ on the measured frequency f_1 is considerably smaller (by a factor of approximately $\beta_0/3$ times) in the pulse method than in the usual methods of measuring critical frequencies [see (20)]. This is connected with the fact that the steady-state charge Q is measured in the circuit of Figure 1 when a current $I_{b0} = I_e/\beta_0$ flows through the base at the time t_1 , while in the ordinary method of measuring, a current $\sim 0.3 I_e$ passes through the base of the critical frequency. This property of the pulse method makes it possible to determine the mean free time of passage of the holes through the base region t_{pr} and, correspondingly, the frequency f_1 "in pure form" without any corrections for the effect of $r_b C_c$. This may find a useful application in technological control in the manufacturing of transistors.

CONCLUSIONS

A basis is provided for the pulse method proposed in [1] for measuring the charge of minority carriers in the base of junction transistors in regard to the highest frequency drift transistors existing at the present time.

In the general case of an arbitrary distribution of the impurity in the base region, a relationship is established between the value of the base charge and the frequency properties of drift transistors.

The possibility of measuring the charge of the order of 10^{-11} k with an accuracy within 10 - 15% is shown.

The influence of the capacitances of the emitter and collector junctions of a transistor is studied. It is shown that the pulse method of measuring the charge makes it possible to determine the following parameters of both alloy-type and drift transistors: the transit time of the minority carriers through the base t_{pr} and, consequently, the frequency $f_1 = 1/2\pi t_{pr}$; the effective lifetime of the minority carriers in the base τ ; the diffusion and static capacitances of the emitter C_{ec} , C_e , and the collector C_{sc} , C_{dc} , C_c .

A number of advantages of the pulse method of determining the parameters of the transistors, compared with the existing measuring methods, is shown.

The author thanks K.S. Rzhavkin sincerely for his valuable advice and constant attention to the work, as well as V.V. Migulin, who made a number of useful comments about the manuscripts of the article.

REFERENCES

1. V.I. Shveykin, Experimental determination of the basic properties of transistors by measuring the charge of the minority carriers of the base. *Radiotekhnika i Elektronika*, 1960, 5, 7, 1158.
2. J. te Winkel, Drift transistor, *Electronic and Radio Engr*, 1959, 36, 8, 280.
3. L.G. Cripps, Transistor high-frequency parameter f_1 , *Electronic and Radio Engr*, 1959, 36, 11, 43.
4. H.L. Armstrong, On calculating the current gain of junction transistors: with arbitrary doping, *IRE Trans. Electron Devices*, 1959, ED-6, 1, 1.
5. Yu.R. Nosov, Junction characteristics of semiconductor diodes. Collection of articles: *Semiconductor Devices and Their Applications*, No. 4, edited by Ya. A. Fedotova, Izd. Sovetskoye Radio, 1958.
6. R.H. Rediker, D.E. Sawyer, Very narrow base diode, *Proc. I.R.E.*, 1957, 45, 7, 944.
7. M.M. Samokhvalov and N.S. Spiridonov, Frequency properties of transistors prepared by the alloying-diffusion method. Collection of articles: *Semiconductor Devices and Their Applications*, No. 3, edited by Ya. A. Fedotov, Izd. Sovetskoye Radio.

Received May 21, 1960

CONVERSION OF A CURRENT PULSE FROM A PHOTOELECTRIC MULTIPLIER INTO A VOLTAGE PULSE BY AN EMITTER FOLLOWER

L.S. Gorn, B.I. Khazanov

This article deals with the characteristics of the conversion of a current pulse from a photoelectric multiplier into a voltage pulse by means of an emitter follower stage. The non-linearity of the conversion is computed and the instability factor is determined. The results of experimental verification are given.

In devices fed by a source with a high internal impedance (a current generator), a linear conversion of a current pulse into a voltage pulse is often required. As a rule, this operation is carried out with a resistance transformer - a cathode follower.

Similar results can be obtained by using an emitter follower. However, it is necessary to take into account the characteristics of the transistors in solving this type of problem. These characteristics include, in the first place, the temperature variation of the transistor parameters and the existence of junction capacitances which are inversely proportional to the applied voltage [1].

Let us consider an emitter follower stage fed by a current generator (Figure 1). The current pulse $i(t)$ is supplied to the divider $R_1 R_2$, which determines the operation of the transistor. The voltage signal is taken from the emitter load R_e . The transfer characteristics of the signal is determined by the capacitances charged by the current generator. The capacitance of the collector junction is noted by C_T , the capacitance of the circuit wiring by C_0 , and the total capacitance load by $C_0 + C_T = C$.

We shall denote the equivalent resistance of the base circuit by R ($\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{\beta R_e}$), the constant voltage across the base-collector junction in the absence of the signal by U , and the input signal by v_{in} .

The output voltage of the emitter follower is $v_{out} = K v_{in}$, where K is the transfer constant. In the general case, this constant depends on the frequency characteristics of the transistor, the frequency spectrum of the transmitted signal, and the amplitude of the signal.

We shall consider the most interesting case for practical applications - the case where the signal builds up during a time equal to a fraction of a microsecond. This corresponds to a spectrum in which the upper frequencies are of the order of 1 - 1.5 Mc, and the ratio of the transmitted frequency to the cutoff frequency of the triode $b = f/f_{\alpha 0}$ does not exceed unity even for alloy-type transistors. For drift transistors, this ratio is considerably smaller.

If $b \leq 1$ and if the signal is small, ($v_{in} \leq 1v$, corresponding to small change of r_e), the transistor is practically a purely active and linear element, the transfer coefficient is practically constant and close to unity, and $v_{in} \simeq v_{out} = v$.

Indeed, if we use the equivalent conversion circuit of the transistor, we can write the repetition factor as [4]

$$K = \frac{R_e}{R_e + \frac{r_b}{|\beta|} + \frac{r_e}{1+jb}}$$

For actual transistors in circuits $r_e \simeq 25$ ohms, $r_b/|\beta| \simeq 40$ ohms, and $R_e = 4.7$ kilohms

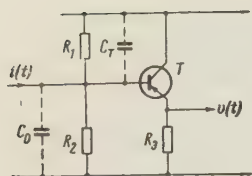


Fig. 1. Circuit diagram of an emitter follower

For the case $b = 1$, the transfer coefficient $K = 0.99$, and the phase shift angle does not exceed a fraction of a degree.

All the following computations are correct in this approximation, i.e., as long as $f_{\alpha 0} \leq (1 \text{ to } 5) 0.3 / \tau_H$.

It is well known that the behavior of the signal $i(t)$ in an RC-circuit of this type can be described by the differential equation

$$\frac{dv}{dt} + \frac{v}{RC} = \frac{i(t)}{C}. \quad (1)$$

For tube circuits, the value of the capacitance C does not depend on the signal formed at the load. This provides a linear conversion of the current into a voltage amplitude [2].

In the case where transistors are used, the value of the capacitance C_T is a function of the voltage across the junction. This results in a nonlinear conversion.

In the general case, $C_T = c(U-v)^{-a}$, where c is a constant for a given transistor, and $a = 1/2$ for alloy-type transistors and $1/3$ for drift transistors [1].

Let us express the value of the capacitance for a signal $C_T(U-v)$ in terms of the capacitance for a stationary state $C_T(U)$:

$$\frac{C_T(U-v)}{C_T(U)} = \frac{U^a}{(U-v)^a}$$

or

$$C_T(U-v) = C_T(U) \left[1 - \frac{v}{U} \right]^{-a}. \quad (2)$$

Then

$$\begin{aligned} C &= C_0 + C_T = C_0 + C_T(U) \left[1 - \frac{v}{U} \right]^{-a}, \\ C &= C_0 \left[1 + \lambda \left(1 - \frac{v}{U} \right)^{-a} \right], \end{aligned} \quad (3)$$

where $\lambda = C_T(U)/C_0$. Substituting Eq. (3) into (1), we obtain

$$\frac{dv}{dt} + \frac{v}{RC_0} \frac{1}{\left[1 + \lambda \left(1 - \frac{v}{U} \right)^{-a} \right]} = \frac{i(t)}{C_0} \frac{1}{\left[1 + \lambda \left(1 - \frac{v}{U} \right)^{-a} \right]}. \quad (4)$$

Equation (4) describes the relationship between the current pulse, the circuit parameters, and the voltage signal at the output. It can be solved in general form by numerical integration. However, if certain simplifications which are permissible for cases of practical interest are made, the equation can be solved analytically.

Let us consider the case of a scintillation counter whose circuit is given in Figure 2. The operation of the emitter follower is determined by a low resistance divider to which the resistance of the base circuit R_1 is connected. This same resistance of a base circuit is the load of a photoelectron multiplier.

A similar circuit is used in recording radioactive radiation, when a phosphor light flash is excited under the action of the ionizing particles. This light flash is converted by a photo-multiplier into a current pulse $i(t)$. The current amplitude i_0 and the total charge flowing through the load R_H is, as a rule, proportional to the energy lost by the ionizing particle in the phosphor. Since this conversion is linear, spectrometric measurement of the nuclear radiation is possible.

As is well known, the shape of the current pulse in the scintillation counter is described by eq. (3)

$$i(t) = i_0 e^{-\frac{t}{\tau}},$$

where τ is the luminescence time of the phosphor (for sodium iodide, which is used most often as a scintillator, $\tau = 0.3 \cdot 10^{-6}$ sec).

To make sure that the complete charge will be corrected on the load capacitance, time constant RC is made considerably larger than the time τ , i.e., the case of practical interest is that where $R \rightarrow \infty$ in eq. (4). The equation is then simplified and assumes the form

$$\frac{dv}{dt} = \frac{i(t)}{C_0} \frac{1}{\left[1 + \lambda \left(1 - \frac{v}{U}\right)^{-a}\right]}. \quad (5)$$

Separating the variables and integrating, we obtain

$$\int_0^{v_m} \left[1 + \lambda \left(1 - \frac{v}{U}\right)^{-a}\right] dv = \frac{1}{C_0} \int_0^{\infty} i(t) dt = \frac{Q}{C_0},$$

where v_m is the voltage amplitude of the signal at the load. Moreover,

$$\begin{aligned} \frac{Q}{C_0} &= v_m + \lambda \int_0^{v_m} \left(1 - \frac{v}{U}\right)^{-a} dv = v_m + \frac{\lambda U}{1-a} \left[1 - \left(1 - \frac{v_m}{U}\right)^{1-a}\right], \\ Q &= C_0 v_m \left\{1 + \frac{\lambda}{1-a} \frac{U}{v_m} \left[1 - \left(1 - \frac{v_m}{U}\right)^{1-a}\right]\right\}. \end{aligned} \quad (6)$$

This equation relates to the voltage amplitude to the charge. For tube devices in which $a = 0$, $Q = v_m(C_0 + C_{in})$. Thus there is a linear conversion of the current pulse into voltage. For transistors, $a \neq 0$, and the conversion is characterized by a certain nonlinearity. The amount of nonlinearity η can be determined in the following way (for small values of the nonlinearity):

$$\eta \simeq 1 - \frac{Q}{v_m(C_0 + C_T)} = \frac{\lambda}{1+\lambda} \left\{1 - \frac{U}{v_m} \frac{1}{1-a} \left[1 - \left(1 - \frac{v_m}{U}\right)^{1-a}\right]\right\}. \quad (7)$$

Hence, it is seen that the nonlinearity depends on the ratio between the signal and the constant voltage across the junction v_m/U , the ratio between the transistor capacitance and the capacitance C_0 , i.e., and the value of λ , and also in the type of transistor (the parameter a).

Let us consider the case of small signals, i.e., $v_m/U \ll 1$.

Using series expansion, we obtain the following equation for the nonlinearity

$$\eta \simeq \frac{\lambda}{1+\lambda} \frac{a}{2} \frac{v_m}{U} = \frac{C_T}{C_0 + C_T} \frac{a}{2} \frac{v_m}{U}. \quad (8)$$

From Eq. (8), it follows that for drift transistors the value of the nonlinearity is smaller than for alloy-type transistors. This is the result of the smaller value of a and, in particular, the small value of capacitance C_T in comparison to the capacitance of the photoelectron multiplier and the circuit C.

Calculations of the value of the nonlinearity of mass-produced transistors may be found interesting. The table gives the values C_T for a number of alloy-type and drift transistors, ($U = 5$ v).

Alloy-type transistors				Drift transistors			
P13, P14, P15		P19		P401		P402, P403	
$C_T, \mu\mu f$	a	$C_T, \mu\mu f$	a	$C_T, \mu\mu f$	a	$C_T, \mu\mu f$	a
50	0.5	12	0.5	15	0.33	10	0.33

For $C_0 = 20 \mu\mu f$, the conversion remains linear within 1% for the following maximum signal amplitudes: $v_{m\max} = 0.3$ v for P13 - P15 transistors; $v_{m\max} = 0.5 - 0.64$ for P401 and P19 transistors and $v_{m\max} = 1$ v for P402 and P403 transistors.

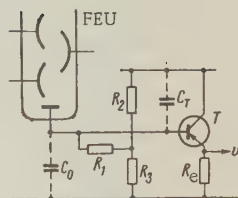


Fig. 2. Circuit diagram of a scintillation counter

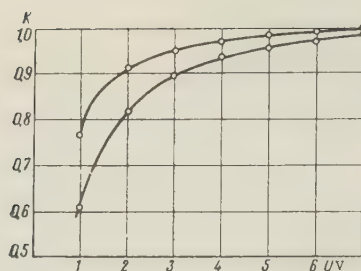


Fig. 3. The signal amplitude as a function of the junction voltage.

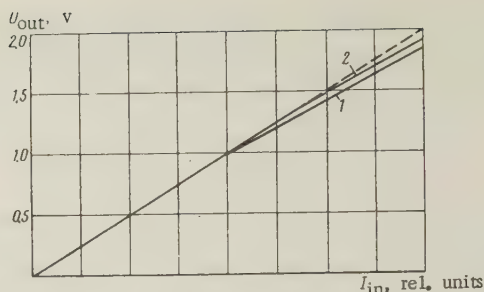


Fig. 4. The linearity of the conversion of the current pulse into a voltage pulse: 1) the P15 transistor; 2) the P403 transistor

It is obvious that the nonlinearity will increase as the supply voltages decrease and as the voltage across the base - collector junction is decreased.

Besides leading to a nonlinear conversion, the dependence of the junction capacitance on voltage, as is quantitatively expressed by Eq. (3), results in a dependence of the voltage pulse at the output of the follower (for a constant current pulse) on the value of the voltage across the junction U .

Let us evaluate the relative magnitude of the change of the signal change dv_m/v_m as the voltage across the junction U varies for the case of small signals:

$$v_m \simeq \frac{Q}{C_0 + C_T}, \quad \frac{dv_m}{dU} \simeq - \frac{Q}{(C_T + C_0)^2} \frac{dC_T}{dU} = v_m \frac{1}{(C_T + C_0)} \frac{C_T}{U} a;$$

hence,

$$\frac{dv_m}{v_m} = \frac{C_T}{C_T + C_0} a \frac{dU}{U} = \frac{\lambda}{1 + \lambda} a \frac{dU}{U}. \quad (9)$$

From eq. (9), it is seen that the instability of the transfer coefficient is related to the instability of the voltage across the transistor. The proportionality factor $C_T/(C_T + C_0)$ usually varies within the limits of 0.1 - 0.3 depending on the type of transistor.

Thus, in constructing accurate circuits, it is necessary to stabilize the supply of the collector. This is the essential difference between transistor and tube-type cathode followers.

The next important distinction between transistors and tubes is the considerable change in their parameters when the temperature of the surrounding medium changes. This also effects the stability of the conversion of the current pulse into a voltage pulse by an emitter follower.

It is known [1] that the change in the voltage across the junction is equal to $\Delta U \simeq \Delta I_{k0} R_b$. Consequently,

$$\frac{dv_m}{v_m} \simeq - \frac{C_T}{C_T + C_0} a \frac{dI_{k0}}{U I_{k0} R_b}.$$

For a 1% change in the signal, the quantity $\frac{dI_{k0}}{I_{k0} R_b}$ can be 0.03 - 0.1. Considering that a temperature of 50 - 60°C, the current I_{k0} can attain values of the order of 50 μ a (i.e., $dI_{k0} \simeq 5 \cdot 10^{-5}$ amp), while $U \simeq 5$ v, we obtain a maximum value of $R_b = 10$ kilohms.

With such a small load, the charge is not completely corrected. This is undesirable. Complete collection of the charge can be attained either by increasing the capacitance C_0 , which leads to a considerable decrease of the signal amplitude at the output, or by increasing the input resistance of the signal by using feedback coupling while retaining a relatively small dc resistance.

The results obtained above are in good agreement with the results of an experimental verification which are indicated in Figures 3 and 4.

Figure 3 shows the amplitude of the output signal as a function of the voltage across the

junction when the current pulse from the photoelectric multiplier is constant. Figure 4 shows the coefficient of conversion of the current pulse into a voltage pulse as a function of the signal amplitude. The small divergences between the curves and the computed values can be explained by errors in the determination of the values of C_0 and C_T .

REFERENCES

1. R. Shea, Transistors and their applications, by R. Shea. Translated from English, GEI, 1957.
2. W. Elmore and M. Sands, Electrons in Nuclear Physics, Translated from English, IL, 1951.
3. J. Burks, Scintillation Counters. Translated from English, IL 1955.
4. Yu.A. Kamenetskiy, Equivalent circuits of transistors. Collection of articles: Semi-conducting Devices and Their Applications, No. 2, edited by Ya. A. Fedotov, 2nd edition, Sovetskoye Radio, 1957, pp. 78-141.

ELECTRONIC CONTROL OF PULSE DURATION IN L-OSCILLATOR WITH JUNCTION TRANSISTOR

B. S. Mel'nikov

On the basis of the piecewise linear theory of inductive relaxation oscillators the report shows the possibility of electronic control of pulse duration by a factor of 100 and greater. The derived design relations describe the real processes with an accuracy of 10-15%.

INTRODUCTION

An L-oscillator is a relaxation oscillator with one reactive element — a transformer. Owing to its structural simplicity, reliability, small dimensions and high efficiency, the L-oscillator has found wide use as a pulse shaper, generator of linearly varying current and direct-voltage converter (references [2, 3]).

In conjunction with the possibility of electronic control of the duration of the generated pulse, the L-oscillator may be used as a blocking oscillator and a pulse-duration modulator.

1. THEORETICAL DISCUSSION

For analysis of the L-oscillator we shall use the piecewise linear theory of the blocking oscillator, describing its behavior during pulse peaking operation (reference [1]). According to this theory the transistor in pulse peaking operation is described by an equation of state which functionally relates the base current to base voltage and collector current:

$$I_b = \frac{1}{R_{bs}}(U_b - U_{b0} - R_{am} I_k), \quad (1)$$

where $R_{bs} = \partial U_b / \partial I_b$ in saturation operation; $R_{am} = \partial U_b / \partial I_k$ in amplifier operation; U_{b0} is a parameter determined from the linearized characteristics of $I_b = f(U_b, I_k)$.

It must be pointed out that the proposed theory is valid only for that operating condition of the L-oscillator at which the rate of change in base current (as defined by the differential equations of the discussed system) is lower than the rate of recombination of minority

carriers at the collector junction (that is, when the lag of the saturated transistor may be disregarded).

The basic diagram of the L-oscillator is shown in Figure 1.

In preparing the initial equations let us use the equivalent diagram for the base circuit, as shown in Figure 2.

$$\begin{aligned} nI_K &= I_b + I_s, \\ \frac{E-U}{n} &= L \frac{dI_s}{dt}, \\ \frac{E-U}{n} &= R_b I_b + U_b + E_b, \\ U &= I_K R_K, \\ I_b &= \frac{1}{R_{bs}} (U_b - U_{b0} - I_K R_{am}). \end{aligned} \quad (2)$$

Here L is the magnetization induction; R_K is the sum of the load resistance (R_{K0}) and the resistance of the collector winding of the transformer; R_b is resistance of the base winding of the transformer; n is the transformation ratio; E is the collector supply voltage; E_b is the bias voltage.

Solving system (2) for I_K and I_b , we obtain

$$I_K = \frac{E}{R_K} \left(1 - e^{-\frac{t}{T}}\right) + \frac{E - n(U_{b0} + E_b) + nI_0(R_{bs} + R_b)}{R_K + nR_{am} + n^2(R_{bs} + R_b)} e^{-\frac{t}{T}}, \quad (3)$$

$$\begin{aligned} I_b &= \frac{ER_{am} + (U_{b0} + E_b)R_K}{R_K(R_{bs} + R_b)} \left(e^{-\frac{t}{T}} - 1\right) + \\ &+ \frac{n(E - n(U_{b0} + E_b) - I_0(nR_{am} + R_K))}{R_K + nR_{am} + n^2(R_{bs} + R_b)} e^{-\frac{t}{T}}, \end{aligned} \quad (4)$$

where

$$T = L \frac{R_K + nR_{am} + n^2(R_{bs} + R_b)}{R_K(R_{bs} + R_b)};$$

I_0 is the zero current in magnetization induction, equal to the current flowing through the base winding of the transformer at the initial moment of pulse generation. According to [1], duration of the generated pulse may be determined from the expression

$$I_b(\tau) \beta' [I_K(\tau)] = I_K(\tau), \quad (5)$$

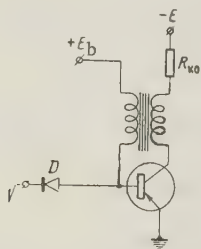


Fig. 1. Basic diagram of L-oscillator

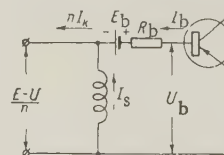


Fig. 2. Equivalent diagram of base circuit of L-oscillator

where β' is the ratio between the collector and base currents at the moment of pulse peak.

For the sake of simplicity let us assume that β' is constant within the operating range of change in collector current (i.e., $\partial\beta'/\partial I_K = 0$). Then simultaneous solution of (3), (4) and (5) leads to the expression for duration of the generated pulse

$$\tau = \frac{LR_{\Sigma}}{R'_b R_k} \ln \left\{ \frac{E \left[\frac{R_{am}}{R_k} + \frac{nR'_b}{R_{\Sigma}} + \frac{nR'_b (R_{am} + nR'_b)}{\beta' R_k R_{\Sigma}} \right]}{E'_b + \frac{E}{R_k} \left(R_{am} + \frac{R'_b}{\beta} \right)} \cdot \frac{E'_b \left[1 - \frac{n^2 R'_b}{R_{\Sigma}} + \frac{nR'_b}{\beta' R_k R_{\Sigma}} \right] - I_0 \left[\frac{R'_b (nR_{am} + R_k)}{R_{\Sigma}} + \frac{R'_b}{\beta' R_k R_{\Sigma}} \right]}{E'_b + \frac{E}{R_k} \left(R_{am} + \frac{R'_b}{\beta} \right)} \right\}, \quad (6)$$

where

$$R_{\Sigma} = R_k + nR_{am} + n^2(R_{bs} + R'_b);$$

$$R'_b = R_{bs} + R_b; \quad E'_b = U_{b0} + E_b.$$

It has been experimentally established that in self-excited operation of the L-oscillator the duty ratio of the pulse is small (1.2-1.5); hence in driven operation with duty ratio greater than 2 it is possible to achieve complete recovery of the circuit, which in our case corresponds to $I_0 = 0$. In addition to this, further simplification of expression (6) will result from the following inequalities, which are suitable for a number of practical cases: $\beta' \gg 1$ (for real transistors $\beta' = 15-20$); $R_k \gg R_{bs} + R_b$; $R_k \gg R_{am}$.

It is not difficult to see that the expression within the braces in (6) is close to unity; hence, we may replace the logarithm with the first term of its power series expansion.

After a simple transformation, the duration of the generated pulse will be expressed by the relation

$$\tau = \frac{L}{R_k} \frac{nE}{U_{b0} + E_b}. \quad (7)$$

Expression (7) shows that the duration of the generated pulse in the first approximation is determined by the parameters of external circuits and is almost independent of the transistor parameters.

The maximum possible modulation of pulse duration is determined by optimum choice of the parameters of the L-oscillator (R_k and n) and the possible change in bias voltage E_b .

The minimum pulse duration corresponds to the maximum permissible voltage at the transistor base $E_{b \max}$ and, since ordinarily $E_{b \max} \gg U_{b0}$,

$$\tau_{\min} = \frac{LnE}{R_k E_{b \max}}. \quad (8)$$

the maximum pulse duration is obtained in the absence of bias voltage ($E_b = 0$) and is determined by circuit parameters R_k , n and E . Inserting $E_b = 0$ into (6) and considering that in actual circuits $\beta' \gg 1$, let us define τ_{\max}/τ_{\min} :

$$\frac{\tau_{\max}}{\tau_{\min}} = \frac{R_{\Sigma} E_{b \max}}{R'_b n E} \ln \frac{E \left(\frac{R_{am}}{R_k} + \frac{nR'_b}{R_{\Sigma}} \right) + U_{b0} \left(1 - \frac{n^2 R'_b}{R_{\Sigma}} \right) - I_0 \frac{R'_b (nR_{am} + R_k)}{R_{\Sigma}}}{U_{b0} + \frac{E}{R_k} R_{am}}. \quad (9)$$

Since n and E are limited, greatest interest attaches to the relation $\tau_{\max}/\tau_{\min} = f(R_k)$.

It is not difficult to see that expression (9) is a function increasing monotonically from R_k and having a limit with $R_k \rightarrow \infty$:

$$\left[\frac{\tau_{\max}}{\tau_{\min}} \right]_{\max} = \frac{(E - nU_{b0}) E_{b \max}}{EU_{b0}}. \quad (10)$$

For actual transistors $E \gg nU_{b0}$; hence

$$\left[\frac{\tau_{\max}}{\tau_{\min}} \right]_{\max} \approx \frac{E_{b\max}}{U_{b0}}. \quad (11)$$

The table presents data indicating the modulation capacity of L-oscillators operating with various transistors.

As seen from the table, in designing L-oscillators with maximum modulation capacity it is necessary to use diffused-junction transistors with high permissible emitter-junction voltage.

In actual circuits the value of τ_{\max}/τ_{\min} is less than given in the table, which is due chiefly to the fact that the collector load resistance R_k may not be greater than the defined value in order to avoid extinction of oscillation.

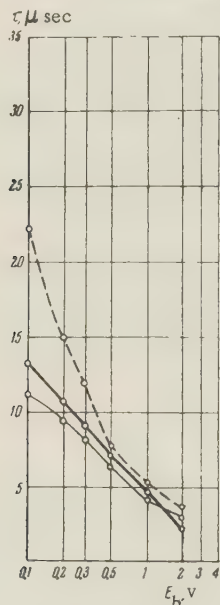


Fig. 3. Modulation characteristic, $R_k = 1$ kilohm

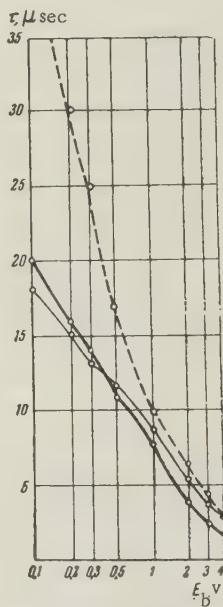


Fig. 4. Modulation characteristic $R_k = 0.5$ kilohm

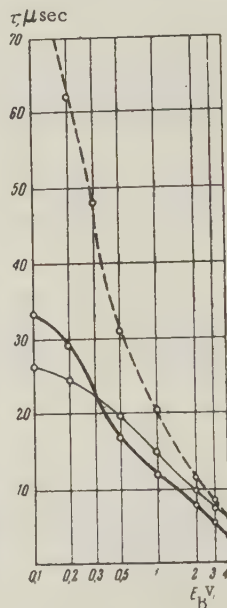


Fig. 5. Modulation characteristic, $R_k = 0.25$ kilohm

Moreover, a decrease in τ_{\max}/τ_{\min} may occur due to incorrect choice of circuit parameters, whereupon the decrease in pulse duration depends on the frequency response of the transistor.

2. EXPERIMENTAL RESULTS

In order to check the theoretical relationships we plotted the modulation characteristics of an L-oscillator using a P14 transistor with parameters $R_{am} = 8.2$ ohms, $R_{bs} = 33.5$ ohms, $U_{b0} = 0.2$ v, $\beta' = 15$.

The circuit values were as follows: $L = 602$ μ h; $E = 10$ v; $R_k = 1, 0.5, 0.25$ kilohm.

The curves in Figures 3, 4 and 5 are presented for comparison of the experimental and theoretical relationships. The theoretical curves were plotted from precise formula (6) (solid lines) and approximate formula (7) (dashed lines). For convenience in comparison the plots were made on a semilog scale. Comparison of the experimental (heavy lines) and theoretical curves show that precise formula (6) for the duration of the generated pulse results in an error of not more than 10-20%, whereas approximate formula (7) gives good correspondence only at sufficiently large bias voltages E_b .

CONCLUSIONS

Electron control of pulse duration is feasible in L-oscillators using diffused-junction transistors.

For ordinary circuits a duration coverage on the order of 10 is attained. For special circuits ($R_k \gg R_{ps}$) using transistors with high permissible emitter-junction voltages a range of duration of 100 and higher may be attained.

With sufficient accuracy for practical purposes, the derived theoretical relations define both the modulation capacity of real L-oscillators and the duration of generated pulses. Hence, the described method of analysis may be used not only for calculation of blocking oscillators and modulators but also for rectangular-pulse shapers, particularly if the latter are driven (wherein the bias voltages are quite large).

REFERENCES

1. B.S. Mel'nikov, Pulse-peak shaping in a blocking oscillator with a junction transistor, *Radiotekhnika i Elektronika*, 1960, 5, 2, 323.
2. Yu.M. Az'yan, G.N. Berestovskiy, L.N. Kaptsov, K.S. Rzhevkin, K.Ya. Senatorov, *Transistors in Regenerative Circuits*, GEI, 1959.
3. *Transistors and Their Applications*, editor, Ya.A. Fedotov, Izd. Sovetskoye Radio, 1960.

Submitted May 3, 1960

STUDY OF THE DEPENDENCE OF THE FREQUENCY OF MASERS ON VARIOUS PARAMETERS. PART II. THE $J=3$, $K=3$ LINE

N.G. Basov, G.N. Strakhovskiy and I.V. Cheremiskin

An experimental study is made of the dependence of the frequency of a maser (the $J=3$, $K=3$ line of $N^{14}H_3$) on the tuning of the cavity resonator, the voltage across the quadrupole capacitor, and the pressure in the source of the molecular beam.

A method is recommended for tuning the frequency of the maser by equalizing the minima (or maxima) on the curves showing the oscillation frequency of the maser as a function of the pressure in the source of the molecular beam. When a one-beam maser is tuned according to this method, its frequency can be reproduced within $\sim 5 \cdot 10^{-10}$.

INTRODUCTION

This article describes experimental studies of the frequency dependence of a maser using the $J=3$, $K=3$ line of ammonia $N^{14}H_3$ on the tuning of the cavity resonator, the voltage across the quadrupole capacitor, and the pressure in the beam source. These studies are necessary for an analysis of the tuning accuracy and of the stability of the maser frequency, which is used as a calibrating frequency, and for a comparison of experiment with theory.

The first studies of the dependence of the frequency and power of the oscillations of a

maser were published in [1, 2, 3]. However the graphs given in these works are insufficiently accurate. This makes it impossible to judge the accuracy which is attained in tuning the masers or to compare theory with experiment. Items of theoretical interest in [4] are the characteristics of masers with cavity resonators in which the E_{011} and E_{012} wave modes are excited. These wave modes make the practical utilization of masers difficult because of the strong influence of the Doppler effect. A comparison of the characteristics which we have obtained with the theory presented in a preceding article [5] shows the good agreement of theory with experiment, and an analysis of these characteristics makes it possible to conduct a fairly detailed analysis of the operation of masers.

1. DESCRIPTION OF THE APPARATUS

A general view of the maser for the 3, 3-line is given in Figure 1. The source of the molecular beam, the quadrupole capacitor, and the resonator were mounted on the same frame, and they were made coaxial by careful adjustment. The source of the beam of molecules is formed by a chamber several tens of cubic centimeters in volume. Ammonia flows into this chamber through a valve allowing continuous regulation. The beam is shaped by a grid having square openings 0.05×0.05 mm in size and a fill factor (transparency) equal to 0.25. The thickness of the grid is 0.05 mm. The diameter of the emerging beam is equal to 6 mm. The pressure of the gas in the beam source was measured with a VT-2 vacuum gauge. An LT-4 electronic pressure-gage tube was connected directly to the chamber of the source (Figure 2).

A diaphragm cooled with liquid nitrogen was placed between the beam source and the quadrupole capacitor. This diaphragm cut out a relatively narrow beam of molecules. To prolong the period of uninterrupted work of the maser, the diaphragm could be rotated to change the aperture, which became clogged with frozen ammonia in approximately 3-5 hours. A source in the form of a single channel 1 - 2 mm in diameter and 10 mm in length was also used. In this case, the maser could also operate satisfactorily without a cooled diaphragm.

The quadrupole capacitor had a length of 150 mm and the gap between the plates was 2 mm. The focused beam of molecules hit a cavity resonator 80 mm long oscillating in the E_{010} mode. The quality factor of the invar cavity resonator, with silver-plated walls on the inside, was 6000 - 8000. The cavity resonator could be retuned over a few megacycles by means of a rod 2 mm in diameter, which could be screwed in through a side wall of the cavity to a depth of 1 mm (leaving the quality factor of the cavity resonator almost unchanged). A 0.1 mm shift of the rod (10 divisions on the dial) changed the resonance frequency of the resonator by approximately 0.5 Mc and shifted the oscillation frequency by 1000 cycles on the average.

The temperature of the resonator was maintained by a thermostat with an accuracy within 0.01°C . A 0.01°C change in temperature changed the frequency of the maser by one cycle. The change in the frequency and power with retuning of the resonator is shown in Figure 3.

The shift of the frequency of the maser as its parameters were changed was made by comparing the frequency with that of a second maser which remained constant within 2 - 5 cycles. Simultaneously, the change in the power was measured in relative units. A block diagram of the setup used for these measurements is shown in Figure 4. According to this system, the frequency of the three masers could be compared two at a time: No. 1 with No. 2, and No. 3 with No. 2. The oscillations of two masers, for instance No. 3 and No. 2, differing by a few hundred cycles from one another, were mixed in a hybrid ring, which serves simultaneously as a good bypass. They then entered a second hybrid ring, which was a balanced mixer, to which the power of a heterodyne klystron 5, tuned to a frequency of 23,830 Mc (i.e., 40 Mc lower than the oscillation

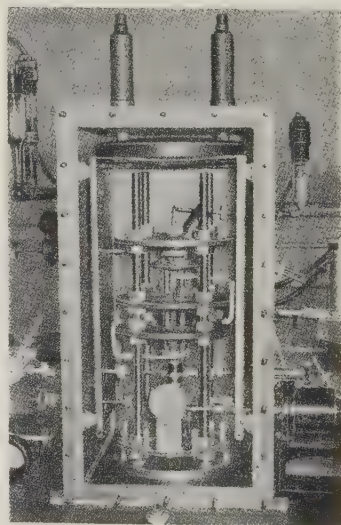


Fig. 1. A maser with a single beam of N^{14}H_3 molecules for the $J = 3, K = 3$ line (the front of the casing is removed)

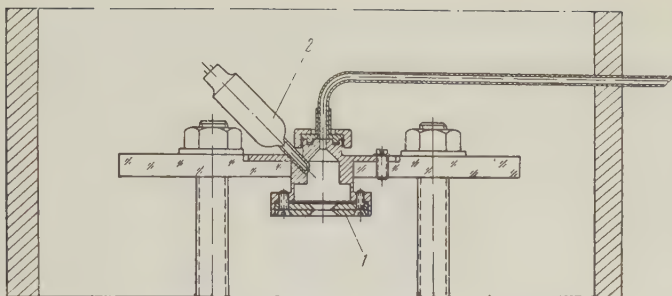


Fig. 2. The source of the molecular beam 1 and the pressure-gage tube 2.

oscillograph. Simultaneously, the signal of a DG-12 audio oscillator was fed to the input of the amplifier of the horizontal deflection of the beam, and the frequency Δf was measured according to the method of Lissajous figures. Simultaneously, the frequency Δf can be measured within one cycle by a frequency meter ICh made of scaling meshes [6], in which the time interval is given by a quartz-crystal oscillator. The relative change in the power was measured simultaneously with the frequency by tapping off part of the power from the cavity resonator and amplifying it in a narrowband IF amplifier (with a 70 kc band) with a double conversion of the frequency. The amplifier was calibrated by means of an auxiliary klystron and a calibrated attenuator. A block diagram of this apparatus is shown in Figure 4.

2. CHARACTERISTICS OF MASERS

With the apparatus described above, we obtained three series of graphs.

1. The amplitude (power W and oscillation frequency Δf as a function of the pressure p in the beam source for different fixed detunings ν of the cavity resonator and for different values of the voltage U of the quadrupole capacitor: $W = f(p)_{\nu, U}$; $\Delta f = f(p)_{\nu, U}$ (Figures 5 and 6).

2. The amplitude (power) and frequency of the maser as a function of the voltage for different fixed pressures and detuning of the resonator: $W = f(U)_{p, \nu}$ and $\Delta f = f(U)_{p, \nu}$ (Figures 7, 8 and 9).

3. The amplitude (power) and oscillation frequency as a function of the tuning of the resonator for fixed values of the pressure and the voltage across the quadrupole capacitor: $W = f(\nu)_{p, U}$ and $\Delta f = f(\nu)_{p, U}$ (Figure 3).

In determining the maser characteristics, we observed the oscillations in the absence of a voltage across the quadrupole capacitor. It seems that this effect is analogous to the "strange" oscillations observed in [2, 3]. Moreover, we noted that, when a large amount of ammonia had frozen in the diaphragm, the maser continued to oscillate if the valve was closed. Apparently, a part of the molecules is evaporated from the diaphragm, sorted by the quadrupole capacitor, and reaches the cavity resonator.

3. COMPARISON OF THE EXPERIMENTAL AND THEORETICAL CHARACTERISTICS

Above we gave the experimental curves for the frequency of the oscillations of the maser as a function of the pressure in the source of the molecular beam, the voltage across the quadrupole capacitor, and the frequency of the cavity resonator. Figures

frequency of the maser) was fed. After the balanced mixer, the power traveled on to a 40 Mc amplifier with a passband of 2 Mc and an amplification factor of 10,000. The frequency of the heterodyne klystron oscillator was stabilized within ~ 50 kc. After a second detector at the output of the IF amplifier, a signal whose frequency was equal to the difference Δf between the oscillation frequencies of masers No. 3 and No. 2 was fed to the input of the EO-7

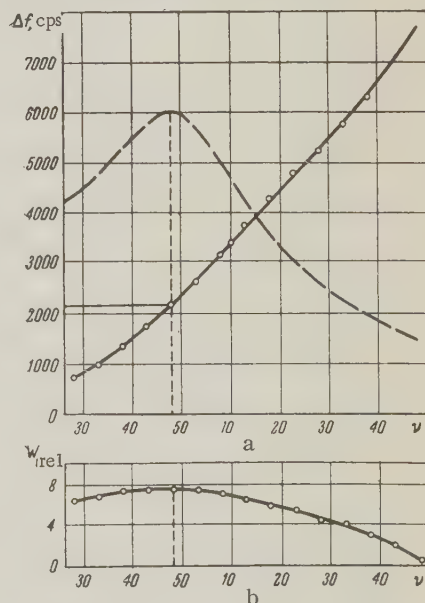


Fig. 3. The frequency (a) and relative power (b) of maser No. 3 as a function of the position of the dial of the tuning screw for the cavity resonator.

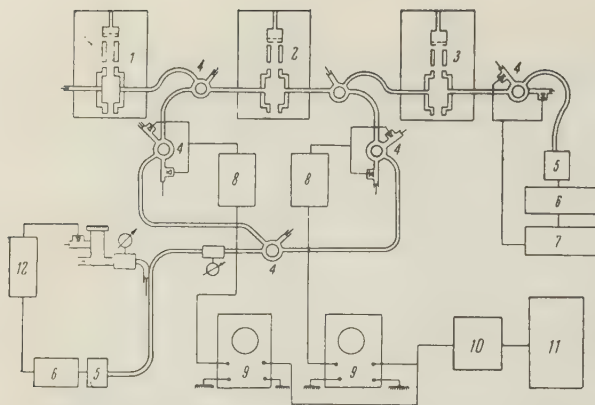


Fig. 4. Block diagram of the system for measuring the frequency of a maser using the $J = 3, K = 3$ line: 1, 2, 3) masers Nos. 1, 2 and 3) 4) hybrid ring-balanced mixers; 5) local klystron oscillator; 6) supply of the klystron; 7) IF amplifier and oscilloscope; 8) IF amplifiers and second detectors; 9) oscilloscope; 10) audio oscilloscopes; 11) frequency meter; 12) 75 kc frequency amplifier and discriminator.

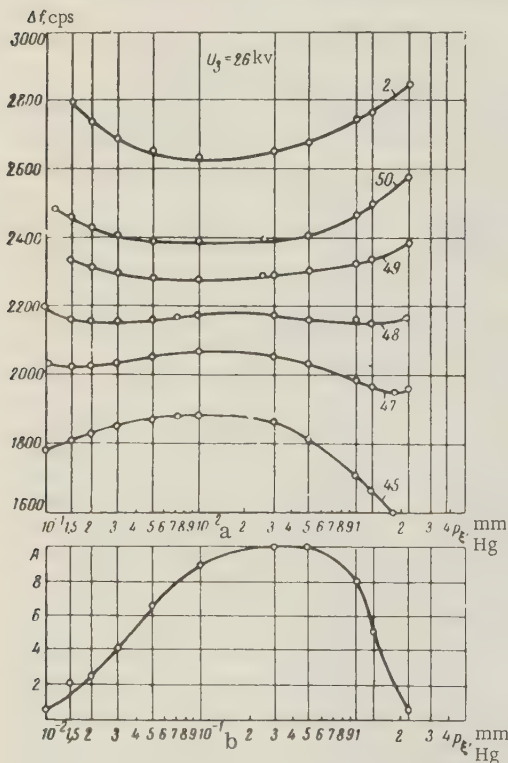


Fig. 5. The frequency (a) and relative power (b) of maser No. 3, as a function of the pressure in the source of the molecular beam for different values of the detuning (the detuning is expressed in terms of divisions of the dial of the tuning screw).

3 and 4 in part I of [5] give curves obtained as a result of theoretical computation. A comparison of the experimental curves with the computed curves shows that they are in qualitative agreement. A quantitative comparison of theory with experiment is made difficult by the following factor: as is pointed out in part I of [5], the theoretical curves are computed as a function of the parameter $\gamma = \frac{d^2 \delta^2}{h^2 \bar{\theta}^2}$, which is pro-

portional to the power of the maser. The power itself is proportional to the pressure only for low pressures in the source of the molecular beam. At large pressures, the power is a complicated function of the pressure, which can hardly be computed theoretically. Experimentally, it is also difficult to make absolute measurements of power with a sufficient degree of accuracy. Therefore, for a qualitative comparison of the theoretical curves with the experimental curves, we measured the relative dependence of the power of the maser, using the $J = 3, K = 3$ line, on

the pressure. We used this relative function to construct theoretical curves. It should be noted, that for small separations between the frequency of the resonator and that of the spectral line, the experimental curves for the frequency are somewhat more complicated functions of pressure than the theoretical curves, when effect III – the irregularities of the radiation of molecules along the resonator (see [5]) – is ignored)

At small detuning, there are clearly four different values of the pressure at which the maser frequency is the same (see Figures 5 and 6). The curve for the maser frequency as a function of pressure when a beam source with a single channel is used has a similar nature. However, the second frequency minimum and the maximum oscillation amplitude are shifted to the region of higher pressures (see Figure 6, a). This is apparently the result of the fact that, for small $(v_r - v_l)/v_l$, when the term $\frac{Q}{Q_1} \frac{v_r - v_l}{v_l} G$ is small, the dependence of the frequency of the maser oscillations is determined by effect III (see [5]).

The maximum change of the frequency due to effect III amounts to 50 cycles when the pressure changes over the entire range from the start to the interruption of the oscillations of a maser with one beam. It should also be noted that the strength of effect III depends on the depth to which the trimming rod of the cavity resonator is inserted. This effect is stronger, the deeper the rod penetrates into the resonator, as can be seen by comparing Figures 5 and 6. The rod penetrates more deeply into the

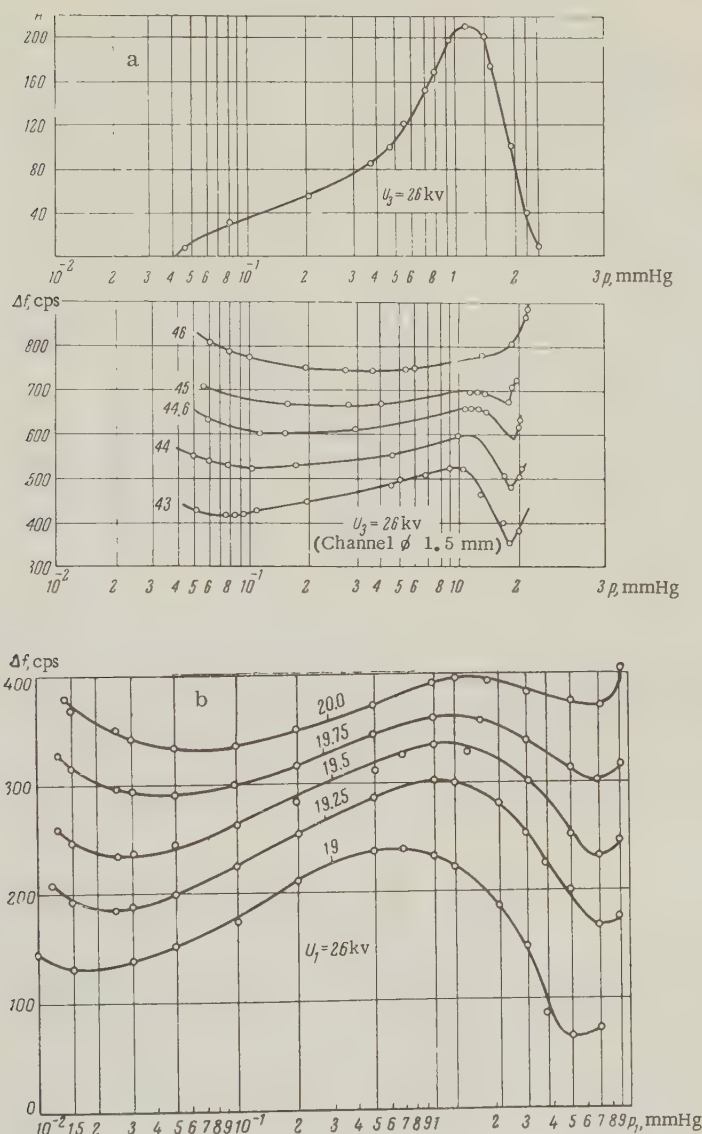


Figure 6

- Frequency and relative power of the maser as a function of pressure when a beam with a single channel is used;
- Frequency of maser No. 1 as a function of the pressure in the beam source. The region of small detunings is studied in greater detail.

cavity resonator of maser No. 1 than into the cavity resonator of maser No. 3.

An examination of the experimental curves illustrating the dependence of the oscillation frequency on the parameters of the maser makes it possible to estimate the maximum possible long term relative stability of oscillations on the $J = 3, K = 3$ line. From Figures 5 and 6, it follows that a change of the pressure from $1.5 \cdot 10^{-2}$ to $5 \cdot 10^{-2}$ mm Hg results in the frequency change of ~ 15 cycles. A change in the voltage across the quadrupole capacitor, (see Figures 7, 8 and 9) by 1 kv results in a frequency change of ~ 5 cycles. When the temperature of the invar cavity resonator changes by 1°C , the frequency of the maser changes by ~ 100 cycles. Thus, if the pressure in the source of the molecular beam is stabilized

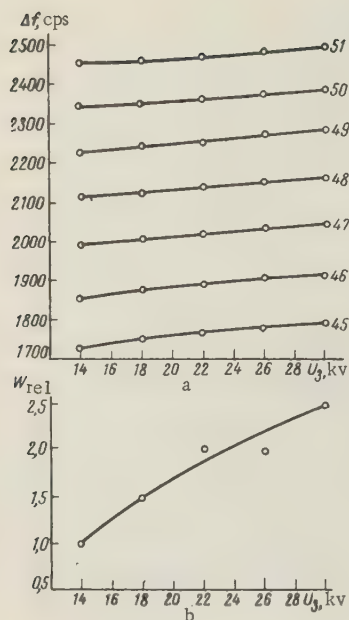
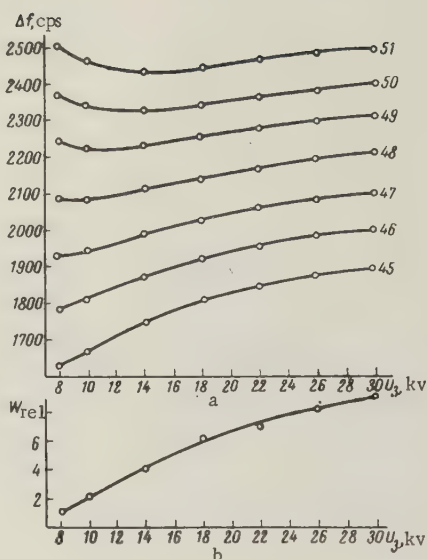


Figure 7. The frequency (a) and relative power (b) of maser No. 3 as a function of voltage across the quadrupole capacitor for different values of detuning, when the pressure in the source of the molecular beam is $p = 1.5 \cdot 10^{-2}$ mm Hg.

Figure 8. The frequency (a) and relative power (b) of maser No. 3 for $p = 1 \cdot 10^{-1}$ mm Hg.



within $\sim 1\%$, the voltage of the quadrupole capacitor within $\sim 0.2\%$, and the temperature of the invar resonator is maintained within $\sim 0.002^\circ\text{C}$ by a thermostat, long term stability of the maser frequency $\Delta f/f \approx 10^{-11}$ can be achieved. To eliminate mutual pulling of the frequency of the two masers used and to prevent the waveguide system from affecting their frequency, it is necessary to use fairly good decoupling.

4. TUNING THE OSCILLATION FREQUENCY OF A MASER BY CHANGING THE PRESSURE

It can be seen from Figures 5, 6, 7, 8 and 9 that the pressure changes the oscillation frequency of the maser more strongly than the voltage does. Therefore, a series of papers [1, 7, 8] proposed using the change of the oscillation frequency of a maser with pressure in tuning the resonator frequency. With this type of tuning, different values of the pressure will correspond to different frequencies of the maser, since the condition $(\partial f / \partial p)_{v, U} = 0$ is not single-valued at small detuning, and the reproducibility of the frequency of a maser can reach typically 10^{-8} . The

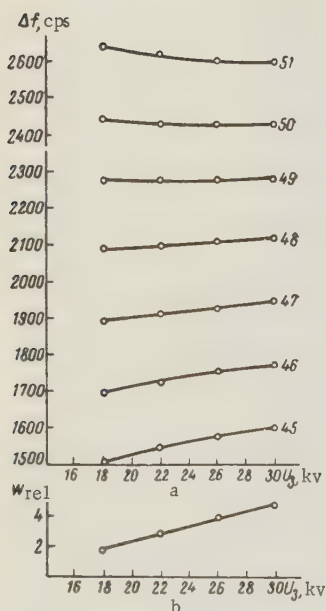


Fig. 9. The frequency (a) and relative power (b) of maser No. 3 for $p = 1$ mm Hg.

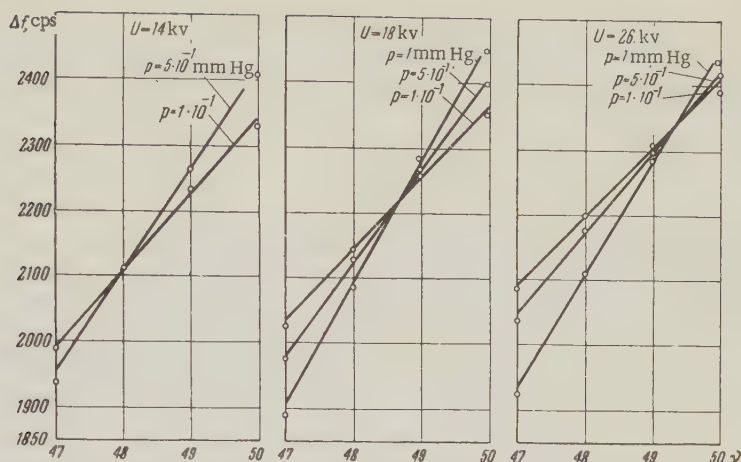


Figure 10. The shift of the intersection point of the curves showing the dependence of the frequency on the detuning at various pressures in the source of the molecular beam when the voltage across the quadrupole capacitor is changed.

tuning frequency also depends strongly on the voltage, as is shown by Figure 10. Let us note that in our design of the maser the working range of the pressures in the source of the molecular beam (see Figures 5 and 6) could be changed from $1 \cdot 10^{-2}$ to 1 mm Hg. Although the oscillation frequency of a maser depends to a smaller extent on changes in the voltage than on changes in the pressure, this effect can also be used to tune the oscillation frequency to the frequency of the spectral line. It follows, from the experimental data (Figures 5, 6, 7, 8, and 9) and the theoretical data, that voltages greater than 20 - 25 kV should be used. Generally speaking, the tuning point of the frequency is not the same when the voltage is changed as when the pressure is changed. This difference can be explained in the following way: from eq. (6) of [5]

$$\nu = \nu_L \left[1 - \frac{Q}{Q_L} \frac{\nu_L - \nu_L}{\nu_L} G(\gamma) : \Delta \right]$$

it is seen that the voltage and the pressure affect the frequency of the spectral line in different ways. While the pressure mostly changes $G(\gamma)$ and scarcely changes Δ (because of the saturation effect), the voltage changes both the first and the second terms fairly strongly. It is difficult to obtain an absolute accuracy for the tuning of a maser better than 10^{-8} by modulating the voltage.

Besides the above methods, it is possible to tune the frequency of a maser to the spectral line by adjusting the minimum (or maximum) of the curve showing the oscillation frequency as a function of the pressure. The experiments show that a single-beam maser using the $J = 3$, $K = 3$ line can be tuned by this method of adjusting the minima of the frequency, as a function of pressure (Figures 5 and 6), with a greater accuracy than when modulation of the pressure or voltage is used. In Figures 6, b, maser No. 1 is tuned in this way for the dial position 19.5, while the minimum on the curve showing the frequency of the maser as a function of pressure with the dial position 19.25 and 19.75 differs by up to 10 cycles. The tuning is made with a fixed voltage across the quadrupole capacitor. In tuning by this method, the reproducibility of the frequency of two masers is $\sim 5 \cdot 10^{-10}$.

An analysis of the theoretical and experimental curves makes it possible to conclude that the use of two opposite beams on the $J = 3$, $K = 3$ line can raise the tuning accuracy of the maser owing to compensation for the effect III. However, the frequency of the spectral line depends on the saturation of the unresolved components of the superfine structure which forms the spectral line (see [5] Δ as a function of γ , Figure 3). In this case as well, the existence of the superfine structure makes it difficult to tune the oscillation frequency with an accuracy greater than $5 \cdot 10^{-10}$. Therefore, to raise the absolute stability of the maser, an ammonia line without superfine structure should be used, for instance the line $J = 3$, $K = 2$ line of $N^{14}H_3$ or a line of $N^{15}H_3$.

REFERENCES

1. J.C. Helmer, J. Appl. Phys., 1957, 28, 212.
2. K. Shimoda, J. Phys. Soc. Japan, 1957, 12, 1006.
3. K. Shimoda, J. Phys. Soc. Japan, 1958, 13, 939.
4. F. Barnes, Proc. I.R.E., 1959, 47, 2085.
5. N.G. Basov, A.V. Nikitin, and A.N. Orayevskiy, Radiotekhnika i Elektronika, 1961, 6, 5, 796.
6. I.D. Murin, Izv. Vuzov. MVO SSSR (Radiotekhnika), 1957, 5, 555.
7. N.G. Basov, Doctor's Dissertation, 1956.
8. K. Shimoda, T.C. Wang, C.H. Tawnes, Phys. Rev., 1956, 102, 1306.

Submitted June 17, 1960

BRIEF COMMUNICATIONS

THE PROBLEM OF STUDIES OF RETARDING SYSTEMS

N. V. Karlov

It is well-known that for effective use of retarding systems it is necessary to use systems in which the energy density of the traveling wave is greatly increased relative to the case of ordinary waveguide propagation. This increase of the energy density is conventionally characterized by the reduction of the group velocity of the traveling wave in the retarding system. The known methods [1, 2] of measuring the group velocity are complicated, and moreover the group velocity alone cannot completely characterize a retarding system. The reason for this is that we are interested in the energy density of the traveling wave, and this quantity depends on the phase velocity, as well as on the group velocity [3].

A direct method of determining the increase of the density of the high-frequency energy in a retarding system is the use of the free radical $\alpha = \alpha$ -diphenyl- β -picryl-hydrazyl (DPPH). It is obvious that the ratio of the value of the line of electronic paramagnetic resonance for a sample of DPPH placed in a retarding system to the value of this line for an identical DPPH specimen placed in the supply feeder is determined by the ratio of the corresponding densities of the high-frequency magnetic energy.

In the case where the high-frequency magnetic field has the same distribution in the feeder and the retarding system, the ratio of the values of the line of the electron paramagnetic resonance determines uniquely the ratio of the group velocities. In this case, if a square-law detector is used, the lines of electronic paramagnetic resonance are inversely proportional to the corresponding group velocity. We should note that the distribution of the magnetic fields in the system can be easily determined experimentally by moving a DPPH specimen about and measuring the signal from the electron paramagnetic resonance line. A knowledge of the distribution of the magnetic fields may be of assistance in determining the distribution of the electric fields.

Besides making it possible to determine the degree of concentration of electromagnetic energy in a retarding system and its distribution in the system, DPPH can be used to determine the configuration of magnetic fields. When DPPH is magnetized, the unpaired electron in it precesses in a plane perpendicular to the magnetizing field. Therefore the intensity of the lines of electron paramagnetic resonance determines the value of the high-frequency magnetic field in this plane with circular polarization in the corresponding direction. This property makes it possible to make a complete analysis of the polarization of the high-frequency magnetic fields of a retarding system by changing the orientation and polarity of the magnetizing field.

It should be noted that direct determination of the configuration and distribution of high-frequency electric fields in a retarding system can be conducted in an analogous manner by using cyclotron resonance in semiconductor crystals.

The above method was used to study the rod delay systems described in [4]. A very large delay of the group velocity in these systems was confirmed. The magnetic fields are concentrated chiefly at the base of the rods, in the plane perpendicular to the rods. The polarization of these fields is nearly circular.

The author is grateful to A. M. Prokhorov for a number of valuable comments.

REFERENCES

1. B. Epstein, G. Mourier, *An. radioelectr.*, 1955, 10, 39, 64.
2. V. B. Shteynshleyger, G. S. Mizezhnikov, *Pribory i Tekhnika Eksperimenta*, 1959, 6, 133.
3. V. M. Kontorovich, *Izv. Vuzov MVO SSSR (Radiofizika)*, 1960, 3, 4, 656.
4. E. G. Solov'yev, E. K. Karlova, *Radiotekhnika i Elektronika*, 1961, 6.

P. N. Lebedev, Institute of Physics,
Academy of Sciences of the USSR

Submitted on November 21, 1960

LETTER TO THE EDITOR

CONCERNING THE ARTICLE "ON A MAGNETRON OSCILLATOR THEORY"

When my article "The Theory of the Magnetron Generator", published in Radio Engineering and Electronics, No. 10, 1960, was in print, I discovered an unfortunate mistake: in Eq. (9), the denominator of the expression under the sign of the Bessel function should contain τ^2 rather than τ . In this connection, some of the subsequent equations Eqs. (10), (11), (12), (12c), (23), (25), (26), (28), (31) and (32) require an additional factor τ in the corresponding places. The presence of this factor makes it possible to simplify considerably Eq. (12) for the electronic conductivity within the first zone of oscillation of the magnetron (for the case $m > 2$). By expanding the trigonometric functions in Eqs. (12b) and (12c) into series, it is possible to obtain

$$Y_e = 2G_0 \frac{J_1(X)}{X} e^{-j(\delta+\pi)},$$

where

$$G_0 = \frac{I_0 M^2 \theta_0 P}{2U_0} (m-2)^2 \sqrt{\frac{m^2 - 2m + 3}{12}};$$

$$X = \frac{M \theta_0 P U}{2U_0} \frac{z}{2} \sqrt{\frac{z^2}{16} + \left(\frac{z}{2} + 1\right) \left(1 - \frac{\tau^2 z^2}{48}\right)} \simeq \frac{M \theta_0 P (m^2 - 4)}{16U_0} U$$

(the notation is the same as in the article. These results and the material of the article may be useful in analyzing the behavior of magnetrons operating close to critical conditions.

V. A. Malyshev
Submitted on January 16, 1960

CHRONICLE

INTERINSTITUTIONAL SEMINAR ON CATHODE ELECTRONICS

EIGHTEENTH SESSION

At the 18th session of the Institutional Seminar on Cathode Electronics, which was held on February 6, 1961, in the Institute of Radio Engineering and Electronics of the Academy of Sciences of the USSR, ten papers were presented.

In their paper entitled "The Conditions for Applicability of the Richardson-Dushman Equation in the Analysis of the Characteristics of a Thermal Electronic Converter", V. A. Grodtko, B. N. Markar'yan, V. S. Zolotarevskiy, and I. M. Rubanovich gave an analysis of the equation for the characteristics of a thermal electron diode converter and showed that the emission coefficients $A = A_0 (1 - \bar{R})$ (where $A_0 = 120.4$ amp/cm² deg⁻², and \bar{R} is the coefficient of reflection) are subject to certain restrictions resulting from the basic laws of thermodynamics. They propose a method for computing the characteristics of a thermal electron converter. This method is based on the use of existing experimental data on the thermal-emission properties of materials, and satisfies the requirements of the laws of thermodynamics.

The authors were asked a number of questions concerning their interpretation of the quantities appearing in the equation for the characteristic of a thermal electron converter. L. N. Dobretsov and I. A. Rezglol', who participated in the discussion, pointed out a number of inaccuracies chiefly connected with the authors' interpretation of the quantity A .

G. V. Stepanov, V. I. Pokalyakin, and M. L. Yelinson presented a paper entitled "The Characteristics of the Emission of Hot Electrons from Intrinsic p - n junctions in SiC Crystals".

The authors observed emission with a high current density from individual luminescent points of small size in pulse operation with different values of the temperature and back bias of the junction. They established that, before the beginning of the process of avalanche multiplication of the carrier, there is a sharp increase in the emission current and a tendency toward its saturation. A spray coating of BaO on the surface of the junction greatly increases the emission current.

The authors answered numerous questions concerning the experimental procedure in their work and the mechanism of the phenomena.

I. M. Bronshteyn and B. S. Frayman presented two papers entitled "Inelastic Scattering of Electrons and Secondary Electronic Emission of Thin Layers of Certain Metals and Semiconductors" and "The Effect of the Work Function on Secondary Electronic Emission".

In the first work, they established experimentally the possible forms of the δ - η diagrams (where δ and η are the slow and rapid components, respectively, of secondary electronic emissions) when one material is placed on another serving as a base. The δ - η diagrams make it possible to determine the cross section δ_0 of the primary electrons penetrating the material and S of the inelastically scattered electrons, and to evaluate the mean free paths of the slow, truly secondary electrons. Results are given for spray coatings of Pb on Si and Al, Ti on Ag, Be and Al; Al on Pb and Ti; and Si on Pb.

In the second work, the authors used the δ - η diagrams to show that, when the work function of an emitter is changed as a result of absorption of extraneous substances on its surface (Ca on Be and Ag; Ba on Be and Ti; Be, Ti, Ag on Ba; Be on Ca), the observed change of the coefficient of secondary emission is caused chiefly by the change of δ . Values of δ_1 , δ_0 , and S are obtained for Ba and Ca, and the role of the "back" flow in the generation of slow, truly secondary electrons is evaluated. The authors were asked many questions concerning the physical explanation of these processes and the interpretation of the experimental function.

In discussing the paper, L. N. Dobretsov pointed out that the value of S used in the work was too low, since the term connected with the change of the spectrum of the electrons moving in the reverse direction as the thickness of the spray coating changed was not taken into account. He recommended that the role of this term be analyzed.

Yu. G. Anikeyev and B. N. Popov presented a paper entitled "The Secondary Emission of Barium Oxide".

The authors measured the secondary emission from BaO with pulsed cathodes with widely varying parameters. These widely varying parameters were the pressure of oxygen (from 10^{-8} to 10^{-5} mm Hg) and the amount of excess barium in the cathode. The absolute value of the secondary-emission coefficient found in the work is in satisfactory agreement with the values obtained by other authors. At temperatures less than 550°C , the coefficient does not depend on T for all active states of the cathode. At high T , the coefficient does not vary with T for cathodes with a very low activity; it increases exponentially for cathodes of medium activity, and decreases negligibly for cathodes of high activity. In the discussion of the paper, A. R. Shul'man noted a number of new interesting and useful results described by the authors, and emphasized that the contradictions in the existing data on the secondary-electron coefficient for BaO can apparently be explained by the processes occurring in the surface layer of BaO when an electric field is applied. He also noted the necessity of studying these processes for a complete understanding of the mechanism of secondary emission from BaO.

V. G. Dmitriyev expressed doubts with regard to the dependence of the coefficient of secondary emission for BaO on T found by the authors in the range of $500 - 600^\circ\text{C}$, and declared that he observed a weak increase of the coefficient for BaO with T in this range of temperatures. The results of the authors can be explained by the presence of oxides of Mo in the BaO layer, stemming from the Mo electrodes used in the experimental apparatus.

F. V. Izmaylov presented a paper entitled "The Theory of Secondary Electron Emission". The author conducted an analysis of the effect of the reflection of primary electrons in the layer of the substance on the emission of secondary electrons. By developing principles used by D. Jonker, the author was able to obtain a more general analytic impression for the secondary emission current.

After answering a number of questions relating to the nature of his laws of the interaction of primary and secondary electrons with the substance, the author indicated that his work was preliminary and that his results could not be regarded as final.

In the discussion of this paper, A. Ya. Vyatskin noted that the work described was extremely interesting and useful, and expressed a wish that the author would use more accurate laws of absorption and retardation in developing it and that he would take into account the angular scattering of electrons in the substance.

E. S. Parilis and P. M. Kishinevskiy presented papers entitled "The Energy Spectrum of Ion-Electron Emission" and the "Mechanism of Ion-Electron Emission and its Dependence on the Ion Velocity".

In the first paper, the authors proposed a mechanism for the exit of excited electrons from a metal. This mechanism makes it possible to explain theoretically the form of the energy spectrum and to compute the position of the maximum, its half-width and the maximum energy of electrons. The emission of an electron in a vacuum is regarded as the result of Auger-recombination of a conduction electron with a hole in the space-charge region formed in the collision of an ion with an atom of the metal. The probability of this Auger-recombination is computed approximately on the basis of Bloch wave functions. A comparison of the resulting theoretical curves with the experimental data is given.

The second paper was a development of the mechanism of kinetic ion-electron emission previously proposed by the authors. This mechanism is based on a statistical examination of the inelastic collision of an ion with a metal atom, accompanied by the formation of a hole in the space-charge zone with subsequent Auger-recombination with a conduction electron, this process leading to the emission of an electron into a vacuum.

The motion of the atoms is considered classically, on the basis of the Thomas-Fermi model. An equation is obtained for the ion-electron emission coefficient γ , giving the dependence of γ on the ion velocity (u). The theoretical curves $\gamma = f(u)$ are compared with the experimental data for the case of different ions in W and Mo. There is good agreement. These papers gave rise to a lively discussion, in which N. N. Petrov, L. N. Dobretsov, S. V. Izmaylov and others participated.

B. Ch. Dyubua and B. N. Popov presented a paper on "The Stability of the Antiemission Properties of Metals under the Action of Oxygen". The authors made computations of the heat of absorption of Ba on the surfaces of different metals. These metals, arranged in a series according to the decrease of the heat of absorption, are Rh, Sr, Pt, Re, Mo, W, Ti, Hf, Zr. Experimental data confirming the computations are given. It was established that Zr had the best antiemission properties of all the metals studied. It was shown that Ti, Zr, and Hf, pure or coated with absorbed Ba, all had higher stability of emission under the action of oxygen than W. The composition of the gases was controlled in the experiment by means of a simple omegatron.

In their paper entitled "Certain Physical Properties of Complex Compounds of Barium and Hafnium Oxide", Ye. S. Zhmud', Ye. P. Ostatschenko, A. I. Figner, and I. V. Yudinskaya described their studies of the physical, chemical and electronic properties of systems prepared from barium oxide and hafnium oxide taken in different ratios. The systems are synthesized by the method of sintering mixtures of the initial powders. The phase composition of samples with different molar ratios of the components was studied by the method of x-ray phase analysis. As a result of the study, the presence of the chemical compound BaHfO was discovered.

R. T. Vikhlyayev, V. A. Gor'kov, A. G. Zhdan

RADIO ENGINEERING AND ELECTRONIC PHYSICS

Institute of Radio Engineering and Electronic Physics,
Academy of Sciences of the USSR

EDITORIAL BOARD

Editor-in-Chief: V.A. Kotel'nikov

Associate Editors: D.V. Zernov, Yu.B. Kobzarev

A.I. Berg	L.N. Dobretsov	A.M. Prokhorov
B.A. Vvedenskiy	A.N. Kazantsev	S.M. Rytov
I.S. Gonorovskiy	S.G. Kalashnikov	V.I. Siforov
V.L. Granovskiy	P.L. Kapitsa	Ya.N. Fel'd
L.A. Zhekulin	V.V. Migulin	S.E. Khaykin
N.D. Devyatkov	A.L. Mikaelyan	B.M. Tsarev
	A.A. Pistol'kors	

Scientific Secretary of Editorial Board: G.A. Bernashevskiy

The English Edition of Radio Engineering and Electronic Physics is mailed to subscribers within 18 weeks after the publication of the original Russian issue.

Russian electronic journals published by the
American Institute of Electrical Engineers
Translated by Royer and Roger, Inc.

	Subscription rates			
	Individuals		Libraries	
	\$	£	\$	£
<i>Radio Engineering and Electronic Physics</i>	28.50	10	57.00	20
<i>Radio Engineering</i>	14.25	5	28.50	10
<i>Telecommunications</i>	14.25	5	28.50	10

Royer and Roger translates and produces
the following Russian scientific journals:

<i>Biophysics</i>	<i>Problems of Oncology</i>
<i>Entomological Review</i>	<i>Radio Engineering</i>
<i>Geochemistry</i>	<i>Radio Engineering and</i>
<i>Geodesy and Cartography</i>	<i>Electronic Physics</i>
<i>Izvestiya, Academy of</i>	<i>Refractories</i>
<i>Sciences of the USSR,</i>	<i>Sechenov Physiological</i>
<i>Geologic Series</i>	<i>Journal of the USSR</i>
<i>Pavlov Journal of Higher</i>	<i>Soil Science</i>
<i>Nervous Activity</i>	<i>Telecommunications</i>

Comments and inquiries regarding *Radio Engineering and Electronic Physics* and other translation journals should be sent to:

International Division
Royer and Roger, Inc.

1000 Vermont Avenue, N.W.
Washington 5, D.C.

41 East 28th Street
New York 16, New York

CONTENTS

N.A. Armand, B.A. Vvedenskiy, Long-Range Tropospheric Propagation of Ultrashort Waves (Survey)	775
N.D. Khaskind, Propagation of Electromagnetic Waves over a Gyrotropic Medium . .	791
A.N. Bruyevich, Asynchronous Oscillations in Self-Excited Oscillator with Two Degrees of Freedom in Overvoltage Operation.	799
B.Ye. Kinber, Decoupling of Two Closely-Spaced Reflectors	809
R.V. Khokhlov, The Theory of Radio Shock Waves in Nonlinear Transmission Lines . .	817
L.S. Benenson, Induced Dipole Impedances in a Waveguide	824
L.V. Kasatkin, Measurements of the Radial Admittance of Axially Symmetrical Systems with Inductive Walls	831
O.A. Barmenkov, S.A. Kornilov, et al., Theoretical and Experimental Investigation of Frequency-Divider Klystrons with Preliminary Bunching.	838
V.I. Kanavets, Harmonics in an Electron Beam at the Output of a Traveling-Wave Tube	848
A.L. Igritskiy, Focusing the Electron Beam in a Traveling-Wave Tube by a Periodic Electrostatic Field.	857
L.E. Bakhrakh, Yu.D. Zharkov, Geometric Parameters of Electron Beams in Ionic Focusing	867
S.B. Gurevich, The Problem of Changing the Signal-to-Noise Ratio Produced by Camera Tubes in a TV System	873
V.M. Breitman, V.S. Kuznetsov, Mathematic Formulation of Problems in Cor-puscular Optics with Space Charge Taken into Account	882
V.I. Shveykin, A Pulse Method of Determining the Parameters of Drift Transistors . .	887
L.S. Gorn, B.I. Khazanov, Conversion of a Current Pulse from a Photoelectric Multiplier into a Voltage Pulse by an Emitter Follower	897
B.S. Mel'nikov, Electronic Control of Pulse Duration in L-Oscillator with Junction Transistor	901
N.G. Basov, G.N. Strakhovskiy, et al. Study of the Dependence of the Frequency of Masers on Various Parameters. Part II. The $J = 3$, $K = 3$ Line	905
BRIEF COMMUNICATIONS	
N.V. Karlov, The Problem of Studies of Retarding Systems	913
LETTER TO THE EDITOR	
V.A. Malyshev, Concerning the Article "On a Magnetron Oscillator Theory" . .	914
CHRONICLE	
Interinstitutional Seminar on Cathode Electronics, Eighteenth Session	914